

社交网络演化计算

——模型、方法与案例

王元卓　于建业
李静远　靳小龙



清华大学出版社

社交网络演化计算
——模型、方法与案例

王元卓　于建业　李静远　靳小龙　著

清华大学出版社
北京

内 容 简 介

本书主要以博弈论、演化博弈论为模型基础，深入地探讨了社交网络演化的建模方法，并结合具体应用场景，介绍了社交网络中的信息、网络和群体的演化计算问题的建模、分析与评价方法，并通过新浪微博、Twitter、Facebook、Google＋、Youtube 等真实数据的演化计算与分析，深入地探讨了社交网络中信息传播、网络结构和群体行为的演化规律，以及如何进行跨网络的时序预测与推荐。

本书既可供计算机、通信、信息等相关专业的教师、研究生和大学高年级学生作为教材或教学参考书，也可供社交网站、电商、网络营销等方面的研究人员和工程技术人员参考。

图书在版编目(CIP)数据

社交网络演化计算：模型、方法与案例/王元卓等著．—北京：清华大学出版社，2018
ISBN 978-7-302-48620-6

Ⅰ．①社… Ⅱ．①王… Ⅲ．①互联网络—数据处理—算法分析 Ⅳ．①TP274

中国版本图书馆 CIP 数据核字(2017)第 258050 号

责任编辑：薛 慧
封面设计：何凤霞
责任校对：王淑云
责任印制：杨 艳

出版发行：清华大学出版社
网 址：http://www.tup.com.cn，http://www.wqbook.com
地 址：北京清华大学学研大厦 A 座 **邮 编**：100084
社 总 机：010-62770175 **邮 购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质量反馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：三河市铭诚印务有限公司
经 销：全国新华书店
开 本：185mm×260mm **印 张**：15 **字 数**：370 千字
版 次：2018 年 3 月第 1 版 **印 次**：2018 年 3 月第1 次印刷
定 价：88.00 元

产品编号：067972-01

前言

随着互联网，尤其是移动互联网的发展，在线社交网络应用也在不断扩展，不仅起到承载信息传播和促进沟通交流的作用，也渗透到购物、支付等人们的日常生活中。社交网络作为对传统人际关系网络在互联网上的一种映射和扩展，渗透到人们网络生活的方方面面，并对人们的社会生活产生深刻的影响。因此，针对社交网络的研究成为当前国内外的研究热点。然而，社交网络的研究，尤其是信息传播、网络结构和群体行为的演化研究，涉及计算机科学、社会学、心理学等多个学科领域，而且社交网络用户规模庞大、网络结构复杂、用户行为多样，导致传统的网络演化分析方法无法适用，迫切需要面向大规模社交网络的新型模型与方法。

社交网络演化计算的一个重要研究方法是从微观上对社交网络的用户进行建模，通过仿真用户间的行为和关系的变化来研究社交网络宏观上的网络结构与中观上的网络群体等的演化规律和内在机理，揭示社交网络中的关系结构、网络群体、网络信息之间的复杂交互关系和互动规律，为社交网络演化分析提供新的认识和理论支撑，并对网络信息的合理利用、网络用户信息行为的正确引导，以及信息网络的有效管理起到重要的指导作用。目前，大多数社交网络建模和演化分析的研究工作是从统计规律的研究展开，侧重于对网络结构的演化分析，缺乏对社交网络中人为因素的考虑，导致对社交网络演化规律的认识不清，缺乏有效的表达与计算。另一方面，网络演化博弈虽然为研究社交网络上用户行为和网络结构的演化规律提供了良好的理论基础，但是现有的网络演化博弈的研究往往侧重于对网络演化现象的理解与理论分析，缺乏对现实社交网络的有效表达。因此，社交网络演化计算是一个重要的研究方向。

我们在基于博弈模型和演化博弈模型的社交网络演化计算领域进行了一系列深入而系统的研究工作，本书主要以博弈论、演化博弈论为模型基础，深入探讨了社交网络演化的建模方法，并结合具体应用场景对社交网络中的信息传播、网络结构和群体行为的演化问题进行建模、分析与评价。书中大部分内容都取材于作者近期在国际、国内一流学术期刊和会议上发表的论文，全面、系统地展示了新的研究成果和进展。

本书内容共 12 章，从结构上可分为 4 个部分。

第一部分是对社交网络与建模方法的基础理论的介绍，包括第 1～4 章。第 1 章从社交网络及其演化的基本概念入手，给出形式化表示和研究要素等基础知识；第 2 章介绍与本书相关的博弈论的基本知识，包括博弈论的常见模型，以及纳什均衡、合作博弈等相关知识；第 3 章进一步给出演化博弈论及网络演化博弈的研究进展，内容上涵盖了演化博弈的基本结构、种群博弈、复制者动态以及演化稳定策略；第 4 章介绍了社交演化博弈模型，包括社交演化博弈模型的建模方法、社交网络个体信息交互行为模型设置方法等。

第二部分是单一网络的社交网络演化计算，包括第 5～7 章。第 5 章分析网络结构对信息传播的影响，包括社交网络信息传播的演化计算模型、方法和实施；第 6 章分析信息行为对网络结构的影响，包括社交网络用户典型信息行为分类，以及用户关系更新策略所造成的网络结构演化；第 7 章给出社交网络群体演化的计算方法，从社交网络群体的评价指标、信息交互群体识别方法、重叠群体相互影响方式、社交网络群体事件的演化分析等方面深入探讨了社交网络群体演化的分析方法。

第三部分是跨网络的社交网络演化计算，包括第 8～10 章。第 8 章介绍跨网络用户偏好的可预测性分析，利用知识库对用户特征进行建模，主要包括显式反馈、隐式反馈和时间维度与空间维度条件下的用户特征建模；第 9 章给出社交网络用户在强关联消费网络中的偏好预测，包括强关联预测模型方法，以及强关联预测的实用性分析；第 10 章探讨社交网络用户在弱关联消费网络中的偏好预测，涉及社交网络与消费网络间的用户账户相似度的计算策略和匹配方法，以及基于用户账户匹配的用户偏好预测模型方法。

第四部分是演化计算工具和实践分析举例，包括第 11 章、第 12 章。第 11 章列举目前常用的博弈仿真计算工具，包括 Gambit、TNG、GAMUT，以及作者为应对社交演化博弈模型仿真开发的 Flock 计算工具；第 12 章给出 4 个基于真实微博数据的社交网络演化计算实例，实际检验本书提出的系列模型和方法在真实社交网络数据上的应用效果。

作者的研究工作得到了国家自然科学基金项目(No. 61173008，No. 91646120，No. 61572473)、国家重点基础研究发展计划(973)项目(No. 2013CB329600，No. 2014CB340400)和北京科技新星项目(No. Z121101002512063)等资金的资助。

本书的一些研究问题和思路得益于 973 项目组的多次研讨，在此对项目组的各位老师表示感谢。北京科技大学研究生陆源参与了本书第 5 章、第 11 章和第 12 章的撰写，中国科学院计算技术研究所研究生刘强参与了第 8 章、第 9 章和第 10 章的撰写，中国科学院计算技术研究所研究生欧陈庚、林谢雄参与了第 5 章的撰写，在此一并表示感谢。

由于作者水平所限，加之社交网络演化计算方法的研究和应用仍处于不断发展和变化之中，书中错误和不足之处在所难免，恳请读者予以指正。

作　者

2017 年 5 月

目录

第一部分　社交网络与建模方法

第三部分 跨网络的社交网络演化计算



社交网络与建模方法

本部分给出本书社交网络及其建模方法的基础理论方法，包括社交网络及其演化的基本知识、博弈论的基础理论、演化博弈的基础理论及当前的研究进展，以及社交演化博弈模型方法。其中，

第 1 章介绍社交网络及其演化的基本概念、形式化表示和研究要素等基本概念和知识。

第 2 章介绍博弈论的基本概念和性质。首先介绍了博弈论的基本框架、理性行为、有限理性等的基础知识，以及常见的博弈模型。随后介绍了博弈论中的重要概念纳什均衡、纳什均衡的存在性以及混合策略纳什均衡。最后介绍了合作博弈的相关知识。

第 3 章介绍演化博弈论及网络演化博弈的研究进展。首先介绍了演化博弈的基本结构、种群博弈、复制者动态以及演化博弈中的重要概念——演化稳定策略。随后在静态网络、动态网络和相互依赖网络这 3 个层面上回顾了近年来网络演化博弈的研究进展。

进一步地，第 4 章给出面向社交网络演化研究的社交演化博弈模型，首先给出了社交演化博弈模型的建模方法，进而根据不同类型的社交网络个体信息交互行为给出了模型设置方法和仿真结果，探讨了个体信息交互行为与社交网络演化的相互影响。

第 1 章 社交网络

1.1 概述

互联网正逐步演变为无处不在的计算平台和信息传播平台。在线社交网站、微博、博客、论坛、维基等社交网络的出现和迅猛发展,使得人类使用互联网的方式产生了深刻变革——由简单信息搜索和网页浏览转向网上社会关系的构建与维护,以及基于社会关系的信息创造、交流和共享。

基于互联网的社交网络正在成为人类社会关系维系和信息传播的重要渠道和载体,对国家安全和社会发展都会产生深远的影响:①社会个体通过各种连接关系在社交网络上构成“关系结构”,包括以各种复杂关系关联而成的虚拟社区;②基于社交网络的关系结构,大量网络个体围绕某个事件而聚合,并相互影响、作用、依赖,从而形成具有共同行为特征的“网络群体”;③基于社交网络关系结构和网络群体,各类“网络信息”得以快速发布并传播扩散形成社会化媒体,并反馈到现实社会,使得社交网络与现实社会之间形成互动,对现实世界产生影响。

针对网络结构、群体和信息之间的互动关系,研究社交网络中各种关系结构(例如,虚拟社区等)的形成机制,发现社交网络关系结构的稳定特性;研究网络群体建模方法,揭示网络群体演化规律,发现网络群体的生成机理;研究网络群体行为规律,分析群体内部的相互作用及群体之间的相互影响,发现社交网络中的信息传播模式;研究多种社交网络之间信息传播的相互影响以及与传统媒体之间的互动规律。这些问题的研究涉及计算机科学、社会学、管理学、心理学等多个学科领域,可以揭示社交网络中的关系结构、网络群体、网络信息之间的复杂交互关系和互动规律,为社交网络分析与网络信息传播研究提供重要的理论支撑。

1.2 社交网络的研究要素

在线社交网络是一种由社会个体集合及个体之间在信息网络上的连接关系构成的社会性结构,包含关系结构、网络群体与网络信息 3 个

要素。其中,社交网络的关系结构是社会个体成员之间通过社会关系结成的网络系统。个体也称为节点,可以是组织、个人、网络 ID 等不同含义的实体或虚拟个体;而个体间的相互关系可以是亲友、动作行为、收发消息等多种关系。其中,社交网络中的虚拟社区为该社交网络的一个子集,且虚拟社区内节点之间关联密切,不同虚拟社区的节点间关联稀疏。网络群体行为是指网络个体就某个事件在某个虚拟空间聚合或集中,相互影响、作用、依赖,有目的地以类似方式进行的行为。基于社交网络的信息传播是指社交网络中的个体与个体之间、个体与群体之间、群体与群体之间的信息传递。这种传递作用可以迭代进行。

与传统的 Web 应用及信息媒体应用相比,社交网络应用具有以下主要的新特点:①信息的发布和接收异常简便、迅速。用户可以通过手机和浏览器随时随地发布和接收信息。②“核裂变”式的信息传播。消息一经发布即刻被系统推送到所有关注者,一旦被其转发,又立即传播到下一批关注者,呈现“核裂变”式的几何级数扩散态势,为普通网民创造了意见扩散的渠道。③人人都有机会成为意见领袖。与传统媒体相比,社交网络服务中的广大网民都有机会通过社交网络形成意见领袖,并在突发事件的产生、发酵、传播、爆炒等环节中具有重要地位。④呈现自媒体形态,能够快速形成虚拟社区。由于社交网络中的个体都具有提供、发布信息的手段和渠道,呈现自媒体形态,并依靠社交网络快速传播信息,快速形成网上的虚拟社区。

1.2.1　网络结构

在社交网络中存在着一定的网络结构,可以用图的形式进行描述和记录。针对社交网络的静态结构特性的分析,可以从宏观、中观和微观 3 个层面展开。

在社交网络的宏观静态结构特性分析方面,大量已有研究验证了在真实世界中各种不同的社交网络具有许多复杂网络所共有的结构特性。例如:六度分隔、小世界现象、无标度、幂律分布和结构鲁棒性等。2003 年,美国哥伦比亚大学的 Watts 等人在 6 万个节点规模的电邮网络上验证了六度分割和小世界模型。2005 年,学者们又通过研究 Yahoo 在线社区的路径长度,发现了网络规模最大时的平均路径长度和有效路径长度分别为 8 和 10,比通常认为的“六度分隔”要大。

在社交网络的中观静态结构特性分析方面,社区结构是研究热点之一。根据社区结构的定义,社区结构可分为不可重叠的社区和可重叠的社区。在不可重叠的社区结构方面,目前最受关注的是 2004 年由美国密歇根大学的 Newman 等人[1]提出的通过寻找使得社区的模块度最大的网络划分来发现网络社区的算法。2006 年,Newman 等人[2]又从矩阵谱分析的角度阐述了模块度优化和传统网络划分的关系,提出了基于模块度矩阵谱分析的社区发现方法。随后,学者们发现,模块度优化方法存在分辨率限制问题,使得基于模块度优化的方法无法识别出一些较小的社区。在可重叠的社区结构方面,2005 年,匈牙利科学院的 Palla 等人[3]提出了一种基于 K-完全子图渗流的重叠社区发现方法,该方法的优点在于能够揭示网络社区间的重叠现象,不足之处是其参数选择缺乏有效的理论指导。2008 年,美国华盛顿大学的 Rosvall 等人[4]把网络社区发现问题看作网络压缩问题,社区发现在于寻找网络的一种高效压缩表示,该压缩表示是保留网络拓扑结构特征和网络压缩比率之间的折中。此后,学者们开始研究网络层次化社区的发现问题。2009 年,中国科学院计算技术研究所的程学旗等人[5]进一步提出了网络层次化重叠社区的发现方法,能够同时揭示网络

的层次化和重叠社区结构，并通过分析信息通路和社区结构的关系，给出了一种基于网络传导率的社区发现方法。

在社交网络的微观静态结构特性分析方面，从微观层面分析节点间的链接关系，发现有向图中节点的出度可以被模型化为一个节点对其他节点的影响程度，即该节点的活跃度；类似地，将有向图中节点的入度考虑为节点受欢迎程度的一个因素，并发现，在社交网络中，流行度同样也服从幂率分布。部分学者还提出了利用频繁模式挖掘算法发现社交网络中微观结构特征。

1.2.2 群体行为

网络用户通过各种信息行为间的作用关系有机地关联在一起，形成具有一定拓扑结构关系的网络群体。一般地，网络群体是指互联网上有相似特质、共同目标或价值观以及团体意识的、由两个(或两个以上的)以一定方式进行线上、线下交互或活动的用户而构成群体。网络群体不但具有一般群体的特点，还具有其自身的特点：依托于社会关系网络，网络群体不受地理等因素的影响，成员分布范围极广，网络群体的时效性比一般群体强等。网络群体形成的方式多种多样，大致归类为以下几种：①由某种固定关系形成的群体，通常是现实社会中的实际关系在网络中的具体映射，例如由同学及校友关系形成的网络校友群体，这类群体的结构比较稳定，形成之后随时间发生的变化不大，成员的活跃程度有限；②由某种不确定的共同需求形成的群体，例如，微信群、微博中的微群、豆瓣的群组等，这类群组中的成员多是由共同的需求或目的聚集在一起，随着时间的变化，群体成员会发生一定的流动，但是流动主要发生在度数相对较低的节点；③由某种突发事件驱动而形成的群体，这类群体具有很强的时效性，通常由一个焦点事件引发而产生，其结构变化与该事件的关注度涨落有关，在事件平息后的短时间内群体就会解散。

在社交网络中，宏观层面的网络群体行为所表现出来的特征比微观层面的个体信息行为更加复杂。网络群体的行为受到社会性和随机性等特性的影响，使得群体的行为往往表现出极大的不确定性，而更加值得关注的现象是群体行为的涌现。由用户行为导致的信息的传播一般都具有从逐渐积累到一夕爆发的特点，这样的积累并不是线性的，也不是用户行为产生效应的简单叠加，而是一系列小的局部变化，每一个小的局部变化都不足以对整体系统造成影响，但当这些变化的积累达到某个临界状态时，整个系统出现临界相变。整个系统表现出来的现象是所有局部行为相互作用的结果。社会关系网络中的大量涌现现象包括信息扩散数量上的涌现、网络结构特征的涌现、特定网络群体的涌现以及特定网络群体行为的涌现等。

为了理论简化和直观表达，对网络群体行为进行分析和建模是社交网络研究的重要内容。目前，对于网络群体行为的模型的研究集中在对网络群体行为定性描述的模型上，主要作为表现个体信息行为复杂过程的一种方法。社会关系网络是一个多维度的复杂系统，网络内部节点之间的相互影响与相互作用频繁。然而，网络群体行为的社会性、随机性和复杂性使得传统的统计研究方法不能对网络群体行为产生的影响进行预测，因此迫切需要新的基础理论和研究方法。目前，网络群体行为模型的相关研究工作主要包括信息传播及观点交互模型、群体行为演化模型、群体结构模型和突发群体事件模型等，但还没有一个公认的

成体系的研究思路。近年来，对社交网络的建模方面进行了大量工作，比如对于社会网络静态社区结构特性的研究、在演化过程中社会网络动态社区结构变化规律的发现等。但是，这些研究工作的关注点主要集中在网络拓扑结构上，对于网络群体行为的预测调控工作还没有有效的研究方法。

1.2.3　网络信息

与传统媒体相比，基于在线社交网络的网络信息传播在传播媒介、传播渠道、信息受众、反馈机制等方面迥然不同。具体来说，在线社交网络体现出以下几个突出的特点。

(1) 平民化、个性化：在在线社交网络中，每个人都从旁观者转变为当事人，每个网民都可以利用网络来表达自己想要表达的观点，传递他们的生活的阴晴圆缺，构建自己的社交网络。在线社交网络因而成为网民张扬个性、表现自我的最佳场所。

(2) 低门槛、易运作：在线社交网络中，用户可以发布文字、音乐、图片、视频等信息，创建属于自己的媒体。这不需要投入多大的成本，也不要求有太多的专业技术知识。其进入门槛低，操作运作简单，让在线社交网络大受欢迎，迅速发展。

(3) 传播迅速、交互性强：在线社交网络大大降低了时空限制，任何时间、任何地点，只要能够上网，用户就可以经营自己的媒体，再加上媒体服务网站所提供的信息推送功能，使得信息能够迅速地传播，时效性大为增强。而且，由于作为一种新的社会媒体形式在线社交网络与受众是零距离的，不但能够迅速地将信息传播到受众中，受众也可以迅速地对信息传播的效果进行反馈，因此其交互性的强大是任何传统媒体望尘莫及的。

因为上述特点，在线社交网络以 Facebook、Twitter、新浪微博和微信等的诞生为标志，在短短十多年的时间里得到了飞速的发展。网络空间给广大网民提供了平等表达自己意见的“新公共领域”。调查显示，有超过一半的网民表示喜欢在互联网上发表评论。网络空间“新公共领域”的特征有助于成为社会冲突的“安全阀”，良好、通畅的对话空间有助于缓解社会矛盾，促进社会和谐。近年来，我国政府积极倡导倡议并引导网络参政议政，广大网民通过互联网渠道评论时事、反映民生、建言献策，网络已经成为推进社会主义民主政治建设的重要力量。

1.3　社交网络的演化

1.3.1　网络信息的传播

信息传播的过程其实也是信息与信息传播者的状态进行演化的过程。目前，得到最广泛且深入研究的信息传播模型是流行病模型。几乎所有的流行病模型都可以用 MSEIR 框架模型来描述，其中，M(passive iMmune)表示被动免疫态，S(Susceptible)表示易感态，E(Exposed)表示潜伏态，I(Infective)表示感染态，R(Recovered)表示恢复态。在流行病模型中，研究最为彻底、应用最为广泛的是 SIS 模型和 SIR 模型[6]。在 SIS 模型中，个体不能对病毒产生免疫，而在 SIR 模型中被感染的个体能够从感染态恢复过来并获得免疫，从而进入恢复态。由于所研究细节的不同，SIS 和 SIR 模型产生了很多的变体，如 SIRS[7]、SIDR[8] 和 SAIR[9]。然而，这些模型都无法反映病毒在使宿主产生传染能力之前有一个潜伏期这

一事实。为解决这一问题,潜伏态被引入了 SIR 模型,从而有了 SEIR 模型[10]。近来,对 SIS 与 SIR 系列模型又有了不少新的改进用以建模信息传播行为。譬如,为了刻画信息传播中广泛存在的点到群的传播模式,清华大学的学者 Yuan 和 Chen 对 SEIR 模型进行了改良,提出了 e-SEIR 模型[11]。在文献[12]中,Hill 等人提出了 SISa 模型来研究行为作为一种特殊的信息通过社会网络在人与人之间的扩散。他们在 SIS 模型中引入了个体自动或自发接受某种行为的可能性。

在信息演化传播模型中,也采用类似于 SI 模型的级联模型和阈值模型。在级联模型中,每当一个个体获得信息,都有一定的概率将信息传播给它的相邻个体。而且个体一旦被感染,将一直处于感染态。研究表明,在均匀网络中,存在有限的传播临界值。当有效传播率大于该临界值时,感染个体能够将病毒传播开,并使得整个网络感染个体总数最终稳定于某一平衡状态;如果有效传播率低于此临界值,则感染个体数呈指数衰减,无法大范围传播。阈值模型则假设每个个体对信息有一个接受阈值,个体之间的关系带有权重,当节点在其所有邻居对其影响之和大于接受阈值时,则接受信息。除了上面几种最流行的信息传播模型之外,学者们还提出了其他一些模型。譬如,在文献[13]中,Young 把社会革新的动力性建模成一个在社交网络上进行的协调博弈游戏。在文献[14]中,Iribarren 和 Moro 发现,信息传播展示出非马尔可夫的分支动力学特性,能够用一个两步 Bellman-Harris 分支过程来建模。

目前对网络信息传播的建模工作通常都基于流行病或类似于流行病模型的级联模型和阈值模型。然而,病毒传播与信息传播有着很显著的差异[15]。例如,在信息传播中,一个个体往往根据多个不同的因素(如信息的内容、自身的喜好等)来决定是否把接收到的信息传播给他的朋友。如果决定要传播的话,则会进一步根据信息的特征、朋友的喜好等因素决定传给哪些朋友,而且往往不是一次就传播给所有要传的朋友。也就是说,现有的信息传播模型往往忽略了传播信息个体之间的差异性,以及他们可以综合多种因素做出决策的能力。因此,现有信息传播模型并不能准确地刻画信息在社会媒体中的传播模式。

1.3.2　网络结构的演化

在社交网络演化分析方面,学者们从社交网络演化中的统计规律展开研究,并提出了面向不同类型社交网络的演化模型。

在社交网络演化规律方面,2005 年美国斯坦福大学的 Leskovec 等人[16]通过演化网络的快照分析,发现了增长的网络顶点与网络直径反关联的现象。2006 年,美国加州大学伯克利分校的 Chakrabarti 等人[17]提出了综合时间点社区质量和时间轴社区变化率的进化聚类算法,建立了社交网络的社区演化模型。2007 年,Chi 等人[18]扩展了相似性计算方法,用图分割(graph cut)作为测度社区结构和社区进化的指标,首次提出进化谱聚类算法。同年,Tantipathananandh 等人[19]根据社会经验列举了社交网络上中观层面子结构变化的各种情况,并构造出相应变化的损耗 cost 评价体系,建立了以个体消耗、组消耗、颜色变化消耗这 3 种因素合成的最优化模型。Leskovec 等人[20]还在 2008 年研究了社交网络的微观演化过程,发现边的生成频度与节点间的已有跳数成反比。

在社交网络演化过程建模方面,为了揭示社交网络的形成和演化机制,学者们提出了众多的演化模型来分析社交网络的形成过程。1999 年,Barabasi 和 Albert[21]提出的 BA 模型则同时考虑了点和边的演化:新加入的节点会引入新边连接到已有的老节点。因此,BA 网

络模型有两个重要特性：一是网络节点不断增加，网络处于动态变化中；二是择优连接，即网络中新增节点连边并不是随机连接到其他节点的，而是与节点的度有关。这两个特点使得 BA 模型显著区别于 ER 模型和 WS 小世界模型。针对权重网络，Barrat 等人[22]在 2004 年提出加权演化网络模型 BBV 模型，该模型认为节点间的权重与网络的拓扑结构是一种共生演化关系，相互驱动。Kumar 等人[23]通过对真实网络数据分析，提出了一种简单的择优连接机制以模拟社交网络的生成过程，发现不同类型用户会对网络生长过程产生不同的影响。Bu 等人[24]提出了一个基于最新更新时间的局部世界模型，用以描述社交网络中的一些结构特征。Yuan 等人[25]针对微博中用户双向好友关系，提出了一种混合连边规则的网络演化模型，该模型生成的网络与真实网络在众多拓扑特征方面能够保持一致。Sun 等人[26]则提出了一种基于社交活动的网络演化模型。

1.3.3 群体行为的演化

社会关系网络中的群体是一个复杂系统，群体由各种具有合作关系的个体有机关联形成，个体间的相互影响和作用较为频繁。对于网络群体行为的建模和预测，已逐渐成为众多学者关注的问题。相关的研究工作包括对网络个体行为的建模与分析、网络群体模型与行为分析、群体行为建模与演化性分析以及群体行为预测等等。但社会关系网络群体中每个个体的行为都体现了对自身效用最大化的追求，而这种追求又往往随着内部因素和外部条件的改变发生着演化，这使得网络群体所表现的行为不能用个体的行为进行简单叠加，对网络群体行为的研究既要考虑群体中的结构关系和信息传播的影响，同时也要考虑人的因素。传统的模型和分析方法无法研究这样的复杂系统，而必须从本质上探索系统化的理论与方法。本项目将演化博弈模型的思路引入网络群体行为的描述中，虽然在复杂网络的研究中已有一些探索，但是目前还没有一个公认的成体系的研究方法和研究思路。

网络个体行为与传统实体行为不同，不仅取决于个人的心理、性格和社会经济地位等因素，还会受到网络信息、虚拟平台互动等元素的影响。因此，对于网络个体行为的研究突破了传统行为学的理论框架，而将视角转向网络系统和信息用户之间的互动模式及其影响因素。

网络个体行为模型 网络个体行为建模主要是通过对人们信息行为动机和结果的刻画，来描述微观层面个体行为的模式和特征。其中，Schramm 等人[27]从传播学的角度，对个体的行为与动机进行了分析，建立了个体行为特征模式。Pei 等人[28]提出利用个体属性中多维性抽取和个体行为特征密切关联的属性取值，从而达到自动学习个体喜好和行为规律的目标。黄志荣等人[29]提出了一种基于关联分析的属性独特性度量方法，用于挖掘群体中的个体行为特征。

个体间观点交互模型 观点互动模型主要研究社会关系系统中由于个体之间决策的影响与外界公共信息的影响，群体中对某些特定事件或事物所持的不同观点的形成(formation)和演化(evolution)等现象，并包括观点的一致性(consensus)与多样性(diversity)保持等问题。Sznajd 模型重点描述的是观点从内向外扩散的现象。该模型最初使用的场景为链状结构[30]，之后，研究者们对其在二维网格以及各种复杂网络结构下的演化进行了研究。Deffuant 模型[31]和 Krause-Hegselmann 模型[32]是有限信任模型的代表，主要描述了具有一定交互门限的个体中的观点演进现象。在 Krause-Hegselmann 模型中，个体选择更新自己观点的折中值不仅受到双方观点值的影响，还受到节点间的权重值的影

响。Di Mare 和 Larota[33]在 2007 年首先提出了观点交互的博弈模型，描述了现实中个体在交互中都倾向于说服对方而避免被对方说服的现象，称为 SO(Stubbornness-Orator)模型。

个体间博弈模型　社会关系网络中，个体的行为通常受自身利益的驱使，即个体倾向于自身效用最大化的方向来选择行为，这就构成了个体间的博弈关系。其中，Santos 和 Pacheco[34]研究了无标度网络上的博弈行为，发现无标度网络结构可以同时使囚徒博弈和雪堆博弈中出现几乎完全合作的状态。Gomez-Gardefies 等人[35]通过研究网络中个体策略的动态组织结构，直观地解释了无标度网络促进合作的原因。Tang 等人[36]研究了网络平均度对博弈中合作行为的影响。Ren 等人[37]发现网络结构的随机性和动力学随机性能够同时促进网络博弈中的合作行为。

根据群体分析的目标和方法，网络群体行为演化分析研究，从网络群体形成的因素、群体的结构、群体的演化等角度进行。Du 等人[38]利用博弈模型研究了不同的伙伴选择方式对个体策略和群体结构的影响，结果发现，个体对声誉的偏好会促使群体形成，并使群体结构呈现星型。Chen 等人[39]提出了群体形成博弈模型，研究了群体发现和群体重叠问题。Palla 等人[3]提出了一种基于完全子图渗流的重叠社区发现方法。目前，完全子图渗流算法仍然是重叠社区发现方法中应用最为广泛的方法之一，被应用到生物网络、信息网络、社会网络等多种类型的网络中。网络群体的演化性是社会关系网络的一个基本特性。近几年，在前述群体发现研究的基础上，群体演化得到了越来越多的关注。Palla 等人[40]基于他们提出的 k-clique 群体发现方法研究群体演化，得到一个有趣的结论，小群体的稳定性是保证其存在的前提，大群体的动态性是其存在的基础。群体内合作的驱动力(driving force)可用演化博弈论(evolutionary game theory)进行建模分析。

合作行为的涌现是指以自我利益最大化为目的的自私人群最终常常会得到合作态占多数的演化结果的现象。单纯针对这一现象的研究主要集中在复杂网络上的演化博弈讨论。Wu 等人[41]研究了动态群体的结构对于合作水平的影响，他们发现，合作个体与非合作个体之间的脆弱连接越多，合作涌现越有可能发生。Poncela 等人[42]在一个不断增长的呈现网络结构的群体中讨论了演化博弈，结果显示，即使是适度的收益优先选择策略也会促使高度数节点的出现；随着网络范围的扩大，出现了高层次的合作，但是作为静态网络相同的网络结构却并没有促进合作。文献[43]提出了一种模型，能够提高在复杂网络中的演化博弈情境下的合作涌现水平。

1.4　社交网络中的分析

1. 网络结构特性与群体行为演化的相互影响

在建立社交网络的演化分析模型时，有一些关键因素会对演化结果产生巨大的影响，这些关键因素也是研究者关注的热点。网络结构特性和个体策略更新机制是对群体行为演化结果有重要影响的因素。

网络结构对群体行为演化的影响　Gomez-Gardenes 等人[44]将群体结构的影响引入了演化博弈模型，得出的结论是网络局部结构提高了合作行为出现的概率，每个群体人数的增加会降低合作水平。Taylor 等人[45]讨论了同质网络结构对合作行为的演化的影响。F. C.

Santos 等人[46]将网络结构的异质特征引入了演化博弈动力学，发现增加异质性有利于合作态的出现。同时，他们阐述了合作行为对无标度网络中的复杂联系的依赖性。

群体行为演化对网络结构的影响　Brian Skyrms[47]的研究指出：随着演化的进行，生成的网络结构倾向于将整个结构分割成一个个小的群体的结论，这激发了很多学者在社区发现方面的研究，比如常见的复杂网络方向[2]，随机游走在近几年也被用来做相关工作[48,49]。

2. 个体行为策略更新机制对群体行为演化的影响

Arne Traulsen[50]研究了允许任意数量的策略在进化过程中突变，发现，当基因突变率高于临界值时，群体由混合策略组成；当低于临界值时，群体趋向同质。他在文献[51]中讨论了个体以一定概率随机选择策略对演化产生的影响，结果证明，当突变率足够小时，策略选择只与收益的排名有关，其他参数的影响可以忽略。William H. Sandholm[52]讨论了个体能够以独立泊松过程更改自身策略的情况，也考虑了噪声对选择的影响。

社交网络表象上的无序性导致了信息生成与传播呈现出一种无序的状态，但是其中群体行为的聚集性和可预测性却使得群体行为具有规律和模式可循。Barabási 等人[53]对匿名手机用户的活动模式的研究发现，人们的活动遵循有规律的模式，93%的人类活动行为是可以预测的，并且人类个体行为的规律性和可预测性在人口统计类别方面没有明显不同。目前，社会关系网络上的用户行为预测的研究主要集中在网络群体中的关系预测和网络群体行为涌现性预测。

3. 社交网络中的个体关系预测分析

Liben-Nowell 和 Kleinberg[54]首次提出关系预测的定义，预测关系在未来时间内是否产生。基于这种关系预测的算法，按方法分为 3 种。一是基于有监督的学习算法，这种方法将预测问题视为分类问题。比较有代表性的工作包括文献[55,56]。尽管有监督的学习算法是目前比较流行的算法，但它在分类特征选择方面经常会面临不平衡性（imbalance），必须依赖于实验数据集的先验分布知识，因而预测结果准确率不高。为了解决这一问题，有学者提出了基于非监督的学习算法的关系预测方法。这类方法试图在两个用户之间定义一个合理的相似度指标，越相似的用户越有可能产生关系。非监督的方法不用考虑关系预测的数据集的先验分布知识，因而准确度较高。经典的相似度指标包括 Adamic-Adar 指标、优先连接（preferential attachment）、Katz 指标。最近人们提出了一些新的相似度指标，如基于地理位置的指标[57]、基于时差的指标[58]、基于用户观点的相似性指标[59]等。第 3 种方法是将关系预测问题视为关系推荐问题[60]，推荐系统中的方法就可以应用到关系预测领域。周涛等人则利用关系预测研究了关系产生的演化机制。

4. 社交网络中的群体行为涌现性分析

涌现性（emergence）是指系统在微观层次的个体相互作用在宏观层次造成的并不直接蕴含在微观相互作用中的现象。涌现性被有的研究者称为是复杂系统研究的圣盘（holy grail）。在网络关系社会中，我们认为，没有全局控制的、各自追求自身利益最大化的各个个体，以相对简单的交互规则相互作用，最终形成宏观上的群体效应（collective behavior）。

涌现性研究最多的成果集中在对于群体中的合作行为和共识的达成过程的讨论上。Villatoro 等人[61]讨论了不同的群体结构和记忆对于社会共识涌现的影响，他们的工作基于 Mukherjee 等人[62]的结论，共识在交互的利益不依赖于之前行为的情况下会发生涌现。Kar[63]使用随机网络模型展现了在某种情况下，群体会涌现出几个共同的稳定平衡点。Yannakakis 等人[64]采用机器学习的方法研究使用最小的代价使得合作在群体中出现涌现。Savarimuthu 等人[65]研究了在交互过程中标签对于合作行为涌现的正面影响。

能够对网络群体行为的涌现性进行预测是人们期望的研究目标，目前，关于这方面的研究并不是很健全。Haglich 等人[66]使用半布尔代数的方法来预测社会网络中的突发行为；文献[67]将自增强聚集模型在机器人环境下模拟实现，发现在该模型下，机器人会在群体决策过程中自动预测结果并选择对自己有利的一方。

小结

本章重点介绍了社交网络研究的三要素——网络结构、群体行为和网络信息，以及三个研究要素的特点、作用和它们之间的互动关系。网络结构的演化、群体行为的演化和网络信息的传播促使了社交网络不断地发生演化。而这三要素也称为社交网络演化研究的入口。针对网络结构、群体行为和网络信息，以及它们之间的互动关系，学者们开展了广泛和深入的研究。相关研究涉及计算机科学、社会学、心理学、管理学等多个学科领域，使得社交网络研究成为多个学科领域、交叉研究研究的热点。本书后续章节将围绕着社交网络研究三要素，从社交网络演化方向入手，介绍社交网络演化分析的模型、计算方法及相关应用。

参考文献

[1] Clauset A，Newman MEJ，Moore C. Finding community structure in very large networks，Physical Review E，2004，70：066111.

[2] Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. Physical Review E，2006，74：036104.

[3] Palla G，Derényi I，Farkas I，Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. Nature，2005，435(7043)：814-818.

[4] Rosvall M，Bergstrom C T. Maps of random walks on complex networks reveal community structure. PNAS，2008，105(4)：1118-1123.

[5] Shen H，Cheng X，Cai K，*et al*. Detect overlapping and hierarchical community structure in networks. Physica A，2009，388(8)：1706-1712.

[6] Newman MEJ. Spread of epidemic disease on networks. Physical Review E，2002，66：016128.

[7] Kuperman M，Abramson G. Small world effect in an epidemiological model. Physical Review Letters，2001，86(13)：2909-2912.

[8] Williamson MM，Leill J. An epidemiological model of virus spread and cleanup. 2003，http://www.hpl.hp.com/techreports/

[9] Piqueira JRC，Navarro BF，Monteiro LHA. Epidemiological models applied to viruses in computer networks. Journal of Computer Science，2005，1(1)：31-34.

[10] Mishra BK, Saini DK. SEIRS epidemic model with delay for transmission of malicious objects in computer network. Applied Mathematics and Computation, 2007, 188(2): 1476-1482.

[11] Yuan H, Chen G. Network virus-epidemic model with the point-to-group information propagation. Applied Mathematics and Computation, 2008, 206(1): 357-367.

[12] Hill AL, Rand DG, Nowak MA, Christakis NA. Infectious disease modeling of social contagion in networks. PLoS Computational Biology, 2010, 6(11): 1-15.

[13] Young HP. The dynamics of social innovation. In: Proc. of the National Academy of Sciences, 2011, 108: 21285-21291.

[14] Iribarren JL, Moro E. Branching dynamics of viral information spreading. Physical Review E, 2011, 84: 046116.

[15] Lv L, Chen D, Zhou T. The small world yields the most effective information spreading. New Journal of Physics, 2011, 13(12).

[16] Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: Densification laws, shrinking diameters and possible explanations. In: SIGKDD. 2005: 177-187.

[17] Chakrabarti D, Kumar R, Andrew T. Evolutionary clustering. In: SIGKDD. 2006. 554-560.

[18] Chi Y, Song X, Zhou D, *et al*. Evolutionary spectral clustering by incorporating temporal smoothness. In: SIGKDD. 2007: 153-162.

[19] Tantipathananandh C, Berger-Wolf T, Kempe D. A framework for community identification in dynamic social networks. In: SIGKDD. 2007: 717-726.

[20] Leskovec J, Backstrom L, Kumar R, *et al*. Microscopic evolution of social networks. In: SIGKDD. 2008: 462-470.

[21] Barabási AL, Albert R. Emergence of scaling in random networks. Science, 1999, 286(5439): 509-512.

[22] Barrat A, Barthélemy M, Vespignani A. Weighted evolving networks: coupling topology and weight dynamics. Physical Review Letters, 2004, 92: 228701.

[23] Kumar R, Novak J, Tomkins A. Structure and evolution of online social networks. In: SIGKDD. 2006: 611-617.

[24] Bu Z, Xia Z, Wang J, *et al*. A last updating evolution model for online social networks. Physica A: Statistical Mechanics and Its Applications, 2013, 392: 2240-2247.

[25] Yuan W, Liu Y. A mixing evolution model for bidirectional microblog user networks. Physica A: Statistical Mechanics and Its Applications, 2015, 432: 167-179.

[26] Sun X, Dong J, Tang R, *et al*. Topological evolution of virtual social networks by modeling social activities. Physica A: Statistical Mechanics and its Applications, 2015, 433: 259-267.

[27] Schramm W. Mass Communications. 2nd ed. Urbana, IL: University of Illinois Press,1960.

[28] Jiang B, Pei J, Lin X, *et al*. Mining preferences from superior and inferior examples. In: Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2008.

[29] Wong RC, Pei J, Fu AWC, *et al*. Online skyline analysis with dynamic preferences on nominal attributes. IEEE Trans. on Knowledge and Data Engineering, 2009, 21(1): 35-49.

[30] Sznajd-Weron K, Sznajd J. Opinion evolution in closed community. Int. J. Mod. Phys. C, 2000, (11): 1157-1165.

[31] Deffuant G, Neau D, Amblard F, *et al*. Mixing beliefs among interacting agents. Advances in Complex Systems, 2000, 3: 87-98.

[32] Hegselmann R, Krause U. Opinion dynamics and bounded confidence models, analysis, and

simulation. Journal of Artificial Societies and Social Simulation (JASSS), 2002, 5(3): 2.

[33] Mare D, Latora A V. Opinion formation models based on game theory. Int'l Journal of Modern Physics C, 2007, 18(9): 1377-1395.

[34] Santos FC, Pacheco JM. Scale-free networks provide a unifying framework for the emergence of cooperation. Phys. Rev. Lett, 2005, 95: 098104.

[35] Gomez-Gardefies J, Campillo M, Floffa LM, *et al*. Dynamical organization of cooperation in complex topologies. Phys. Rev. Lett. 2007, 98: 108103.

[36] Tang CL, Wang WX, Wu X, *et al*. Effects of average degree on cooperation in networked evolutionary game. European Physical Journal B, 2006, 53: 411-415.

[37] Ren J, Wang WX, Qi F. A method to find community structure based on information centrality. Phys. Rev. E, 2004, 70: 056104.

[38] Du F, Fu F. Partner selection shapes the strategic and topological evolution of cooperation. Dynamic Games and Applications, 2011, 1(3): 354-369.

[39] Chen W, Liu Z, Sun X, *et al*. Community detection in social networks through community formation games. In: Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence (IJCAI'11). 2011: 2576-2581.

[40] Palla G, Barabási AL, Vicsek T. Quantifying social group evolution. Nature, 2007, 446(7136): 664-667.

[41] Wu B, Zhou D, Fu F, *et al*. Evolution of cooperation on stochastic dynamical networks, Plos One, 2010, 5: e11187.

[42] Poncela J, Gomez-Gardenes J, Traulsen A, *et al*. Evolutionary game dynamics in a growing structured population. New Journal of Physics, 2009, 11: 1-13.

[43] Zhang J, Zhang C, Chu T, *et al*. Cooperation in evolutionary games on complex networks. In: Proc. of the 2010 49th IEEE Conf. on Decision and Control (CDC). 2010: 1785-1790.

[44] Gomez-Gardenes J, Romance M, Criado R, *et al*. Evolutionary games defined at the network mesoscale: the public goods game. Chaos, 2011, 21.

[45] Taylor PD, Day T, Wild G. Evolution of cooperation in a finite homogeneous graph. Nature, 2007, 447: 469-472.

[46] Santos FC, Pacheco JM, Lenaerts T. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. In: Proc. of the National Academy of Sciences of the United States of America. 2006, 103: 3490-3494.

[47] Skyrms B, Pemantle R. A dynamic model of social network formation. In: Proc. of the National Academy of Sciences of the United States of America. 2000, 97: 9340-9346.

[48] 王元卓,靳小龙,程学旗. 网络大数据:现状与展望. 计算机学报, 2013, 36(6): 1125-1138.

[49] Lai DR, Lu HT, Nardini C. Enhanced modularity-based community detection by random walk network preprocessing. Physical Review E, 2010, 81: 066118.

[50] Traulsen A, Claussen JC, Hauert C. Coevolutionary dynamics in large, but finite populations. Physical Review E, 2006, 74: 011901.

[51] Traulsen A, Hauert C, De Silva H, *et al*. Exploration dynamics in evolutionary games. In: Proc. of the National Academy of Sciences. 2009, 106: 709-712.

[52] Sandholm WH. Stochastic evolutionary game dynamics: foundations, deterministic approximation, and equilibrium selection. In: Sigmund K, ed. Evolutionary Game Dynamics. American Mathematical Society, 2011. 111-141.

[53] Song C，Qu Z，Blumm N，Barabási AL. Limits of predictability in human mobility. Science，2010，327(5948)：1018-1021.

[54] Liben-Nowell D，Kleinberg JM. The link-prediction problem for social networks. In：CIKM'03. 2003：556-559.

[55] Backstrom L，Leskovec J. Supervised random walks：predicting and recommending links in social networks. In：WSDM'11. 2011：635-644.

[56] Lichtenwalter R，Lussier JT，Chawla NV. New perspectives and methods in link prediction. In：KDD'10. 2010：243-252.

[57] Sadilek A，Kautz H，Bigham J. Finding your friends and following them to where you are. In：WSDM'12. 2012：723-732.

[58] Lou T，Tang J，Hopcroft J，*et al*. Learning to predict reciprocity and triadic closure. ACM Trans. on Knowledge Discovery from Data，2010，9(4)：Article 39.

[59] Yin D，Hong L，Davison B. Structural link analysis and prediction in microblogs. In：CIKM'11. 2011：1163-1168.

[60] Deng H，Han J，Li H，*et al*. Exploring and inferring user-user pseudo-friendship for sentiment analysis with heterogeneous networks. In：Proc. of the 2013 SIAM Data Mining Conf.（SDM'13）. Austin，TX，2013.

[61] Villatoro D，Sen S，Sabater-Mir J. Topology and memory effect on convention emergence. In：IEEE/WIC/ACM Int'l Conf. on Web Intelligence and Intelligent Agent Technology Workshops. 2009，155：233-240.

[62] Mukherjee P，Sen S，Airiau S. Norm emergence with biased agents. Int'l Journal of Agent Technologies and Systems（IJATS），2009，1(2)：71-84.

[63] Kar S，Moura JMF. Global emergent behaviors in clouds of agents. In：ICASSP. 2011：5796-5799.

[64] Yannakakis GN，Levine J，Hallam J. Emerging cooperation with minimal effort rewarding over mimicking. IEEE Trans. on Evolutionary Computation，2007，11(3)：382-396.

[65] Savarimuthu S，Purvis M，Purvis M. Emergence of sharing behavior in a multi-agent society using tags. In：IEEE/WIC/ACM Int'l Conf. on Web Intelligence and Intelligent Agent Technology. 2008，310：527-530.

[66] Haglich P，Rouff C，Pullum L. Detecting emergence in social networks. In：IEEE Int'l Conf. on Social Computing. 2010，107：693-696.

[67] Garnier S，Jost C，Jeanson R，*et al*. Collective decision-making by a group of cockroach-like robots. In：Swarm Intelligence Symp. 2005：233-240.

第 2 章 博弈论

2.1 概述

博弈论(game theory)又称对策论,是研究决策主体的行为发生直接相互作用时的决策以及这种决策的均衡问题的学科。它既是应用数学的一个分支,也是运筹学的一个重要学科。

博弈论思想古已有之。我国古代的《孙子兵法》就不仅仅是一部军事著作,而且算是最早的一部博弈论专著。博弈论最初主要研究象棋、桥牌、赌博中的胜负问题,人们对博弈局势的把握只停留在经验上,没有向理论化发展,正式发展成一门学科则是在 20 世纪初。1928 年冯·诺依曼(von Neumann)证明了博弈论的基本原理。1944 年,冯·诺依曼和经济学家摩根斯坦(Morgenstern)共著的划时代巨著《博弈论与经济行为》将 2 人博弈推广到 n 人博弈结构,并将博弈论系统地应用于经济领域,从而奠定了这一学科的基础和理论体系,标志着博弈论的创立。小约翰·福布斯·纳什(John Forbes Nash Jr.)的开创性论文《n 人博弈的均衡点》(1950 年)和《非合作博弈》(1951 年)利用不动点定理证明了均衡点的存在,明确提出了博弈论中最重要的概念——“纳什均衡(Nash equilibrium)”的概念和均衡存在性定理,为博弈论的一般化奠定了坚实的基础。博弈论作为经济分析的主要工具之一,对产业组织理论、委托代理理论、信息经济学等经济理论的发展做出了非常重要的贡献。1994 年的诺贝尔经济学奖颁发给了纳什等 3 位在博弈论研究中成绩卓著的经济学家。截至 2016 年,共有 7 届诺贝尔经济学奖授予在博弈论研究方面有着重大成就的经济学家。自 20 世纪 90 年代以来,博弈理论在经济、社会和计算机等领域中得到广泛应用。

今天,博弈论已经发展成较为完整的学科。根据不同的基准,博弈有着不同的分类。一般认为,博弈主要可以分为合作博弈和非合作博弈。合作博弈也称为正和博弈,是指博弈双方的利益都有所增加,或者至少是一方的利益增加,而另一方的利益不受损害,因此整体利益是有

所增加的。合作博弈研究人们达成合作时如何分配合作得到的收益,即收益分配问题。合作博弈采用的是一种合作的仿真或是一种妥协。而对于非合作博弈,则是指博弈参与人之间不可能达成具有约束力的协议,所有参与人的行动都当成个体行动。它强调个体进行自主决策,研究人们在利益相互影响的局势中如何选择决策以使自己的收益最大,即策略选择问题。合作博弈与非合作博弈之间的区别在于发生相互作用的博弈参与人之间有没有一个具有约束力的协议,如果有,博弈是合作博弈,反之,则是非合作博弈。例如,有两个寡头企业,如果它们之间达成一个协议,联合最大化垄断利润,并且各自按这个协议生产,就是合作博弈。它们面临的问题是如何分享合作成功带来的剩余价值。而如果这两个企业之间的协议不具有约束力,每个企业都只选择自己的最优产量(或价格),那么就是非合作博弈。此外,合作博弈强调的是团体理性(collective rationality),强调的是效率、公正、公平。而非合作博弈强调的是个人理性、个人最优决策,其结果可能是有效率的,也可能是无效率的。

从行为的时间序列性角度,博弈论分为两类:静态博弈和动态博弈。静态博弈是指参与人同时采取行动,或者尽管有先后顺序,但后行动者不知道先行动者的策略;动态博弈是指在博弈中,参与人的行动有先后顺序,且后行动者能够观察到先行动者所选择的行动。例如,"石头,剪刀,布"的游戏是同时决策的,属于静态博弈,而棋牌类游戏等决策或行动是有先后次序的,则属于动态博弈。

按照参与人对其他参与人的了解程度,博弈又可以分为完全信息博弈和不完全信息博弈。完全信息博弈是指在博弈过程中,每一位参与人对其他参与人的特征、策略空间及收益函数有准确的信息。如果参与人对其他参与人的特征、策略空间及收益函数信息了解得不够准确,或者不是对所有参与人的特征、策略空间及收益函数都有准确的信息,则在这种情况下进行的博弈就是不完全信息博弈。

目前经济学家们所谈到的博弈论一般是指非合作博弈,这是因为合作博弈论比非合作博弈论更复杂,理论上的成熟度远远不如非合作博弈论。非合作博弈可分为:完全信息静态博弈、完全信息动态博弈、不完全信息静态博弈和不完全信息动态博弈。与上述4种博弈相对应的均衡概念为:纳什均衡、子博弈精炼纳什均衡、贝叶斯纳什均衡和精炼贝叶斯纳什均衡(perfect Bayesian Nash equilibrium)。

博弈论的研究方法与其他研究社会经济现象的学科一样,都是从复杂的现象中抽象出基本的元素,对这些元素构成的数学模型进行分析,而后逐步引入对其形势产生影响的其他因素,从而分析其结果。从理论上讲,博弈论是研究理性的行动者相互作用的形式理论,而实际上已深入到经济学、社会学、计算机科学等领域而被广泛应用。

2.2 博弈论基础

2.2.1 博弈论的基本构架

本节将介绍博弈论的基本构架,主要涉及博弈的构成要素、博弈的表示方法以及博弈的分类等。

1. 博弈的构成要素

博弈论被定义为是对智能的理性决策者之间冲突与合作的数学模型的研究,博弈论为

分析多个参与者相互影响的决策理论提供了数学方法[1]。

博弈的元素有局中人、信息、策略和支付。博弈过程描述为局中人根据可用信息选择自己的策略,该策略产生一个支付。

定义 2.1(局中人)　博弈所涉及的参与人称为局中人,局中人是理性的和智能的。

定义 2.2(策略)　策略是一个完全可能发生的计划或者决策规则,它定义局中人对博弈空间的每种状态所选择的行动。局中人 i 的所有策略集合以 C_i 表示,所有局中人可以选择的策略组合为 $C=\prod_{i\in N} C_i$(N 是局中人集合)。任一局中人 i 的一个随机策略都是 C_i 上的一个概率分布,用 $\sigma_i(c_i)$ 表示随机策略,则局中人按照随机策略组合 σ 独立地选择他们的策略,那么选择策略 C_i 的概率是 $\prod_{i\in N}\sigma_i(C_i)$ 。

定义 2.3(支付)　博弈论中的支付是指"收益",常以效用尺度来度量,确定在不同的策略组合下博弈结束时各局中人的收益。用 $u_i(c)$ 表示局中人 i 在博弈中得到的期望效用支付。

2. 博弈的表示方法

博弈有两种基本的表示方法:扩展式(extensive form)与规范式(normal form)。它们将局中人、规则、结局和效用以简明且直观的形式表示出来,极大地方便了对博弈的分析。

博弈的扩展式是用树图对一个博弈的基本构成要素,包括局中人、可选行动与次序、结局以及效用,进行的描述。博弈树在几何上由点和线段构成。博弈树上的点称为决策点,每个决策点代表相应局中人的一个决策场景。在每个决策场景下,局中人需要决定下一步的行动选择。而博弈树上与各决策点相连的线段则代表供各局中人选择的行动。决策点上方的线段代表能导致此决策场景的行动,而决策点下方连接的线段则代表此决策场景下局中人能够选择的所有行动。决策点的前后次序描绘了整个博弈的行动次序。一般博弈的第 1 个决策点被画为空心圆,代表起始决策点,其他决策点均为实心圆,最后的决策点则代表博弈行为的结束,其后紧跟着的将是一个结局的效用。图 2-1 是对"石头,剪子,布"游戏的扩展式描述。甲从起始决策点(空心)出发,有 3 种可供采取的行动。对应甲的任意一种行动,乙从自己的决策点出发也有 3 种选择。最后括号中分别是不同结局下甲和乙的效用。

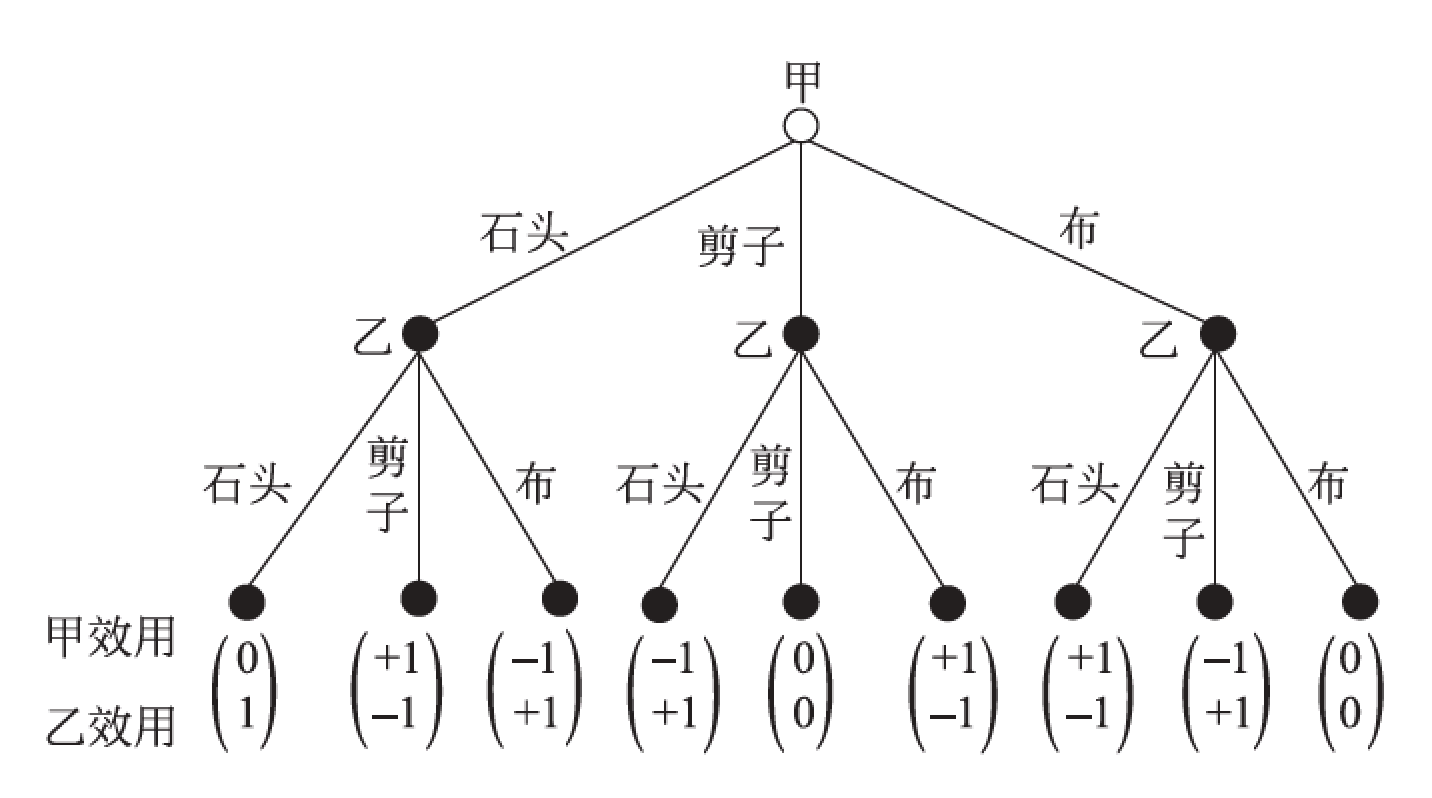


图 2-1　"石头,剪子,布"游戏的扩展式

博弈的扩展式能够直观地描述博弈进程，有助于策略这一概念的理解。可以用一种更加简单的方式，即规范式，来描述博弈。博弈的规范式又称为策略式(strategic form)，它直接用一个博弈中所有可能的策略集合来描述该博弈。规范式可以看成扩展式的浓缩版本。在有 N 个局中人的博弈中，规范式包括：①N 个博弈中所有可能的纯策略组合 S，即每个局中人纯策略集合的笛卡儿乘积 $S_1 \times S_2 \times \cdots \times S_N$。②每种纯策略组合带给各个局中人的效用，用效用函数 $u_i: S \rightarrow \mathbb{R}$ 来表示，$i=1,2,\cdots,N$。"石头，剪子，布"游戏的规范式包括：纯策略组合{石头，剪子，布}×{石头，剪子，布}以及甲和乙的效用：

$$u = \begin{cases} 1, & \text{如果石头对剪子、剪子对布或布对石头} \\ -1, & \text{如果与以上相反} \\ 0, & \text{如果策略相同} \end{cases}$$

以上描述仍显得繁琐，对于只有两个局中人的博弈，其规范式还可以用一种更为紧凑的矩阵来表示。"石头，剪子，布"游戏的规范式见表 2-1，横向为甲的 3 种策略，纵向为乙的 3 种策略，9 种策略组合对应 9 个方格。在每个方格中，左边的实数是乙在该策略组合下的效用，右边的则是甲的。

表 2-1　"石头，剪子，布"游戏的规范式

乙 \ 甲	石头	剪子	布
石头	0,0	1,0	−1,0
剪子	−1,0	0,0	1,0
布	1,0	−1,0	0,0

3. 博弈的分类

对博弈的分类，一个很自然的标准就是其进行的步数。根据其步数的不同，博弈可以分为一次同时博弈(one-shot simultaneous game)和动态博弈(dynamic game)。在一次同时博弈中，所有的局中人同时采取各自的一种行动，比如一次猜拳。这种博弈也称为静态博弈，因为博弈过程没有来回。而在动态博弈中，局中人采取的行动则具有先后次序，比如下棋。博弈的另一个分类标准是看局中人的类型是否为对手所知。局中人的不同类型决定了其在相同结局下的不同效用。比如，两个企业的成本(类型)不同，在赢得同一个项目时的利润(最终效用)也会不同。如果局中人的类型是私人信息，则称这种博弈为不完全信息博弈；如果局中人类型为所有人共知，则这种博弈就是完全信息博弈。

2.2.2　理性行为

通常假定每个局中人都是"理性的"，这种理想是建立在这样的意义之上的，即局中人知道其选择内容，对未知的事物形成预期，具有明确的偏好，并在经过一些优化过程后审慎地选择其行动。排除不确定性因素后，下面的要素便组成了一个理性选择模型：

- 一个行动(action)集合 A，局中人从 A 里做一个选择；
- 一个上述行为的可能结果(consequence)集合 C；
- 一个结果函数(consequence function) $g: A \rightarrow C$，g 使每个行动与一个结果相对应；

- 一个集合 C 上的偏好关系(preference relation)(一个完全的、可传递性的、自反的和二元的关系)。

有时专门给定一个效用函数(utility function)$U: C \rightarrow R$ 来表示局中人偏好。效用函数确定了一个偏好关系$\geqslant$,$x \geqslant y$ 当且仅当 $U(x) \geqslant U(y)$。满足下列条件:

给定任何一个集合 $B \subseteq A$,A 为某个特别情形下的可能行动集合。一个理想局中人选择一个可能行动a^* ($a^* \in B$)。若a^* 对所有 $a \in B$ 满足 $g(a^*) \geqslant g(a)$,则a^* 为最优的。相应地,它解决了问题$\max_{a \in B} U(g(a))$。值得注意的是,使用这个决策模型要假定每个局中人在不同集合 B 选择时使用同一个偏好关系。

局中人通常要在不确定条件下进行决策,其可能:

(1) 不能确定环境的客观参数;

(2) 对博弈中发生的事件不很清楚;

(3) 不能确定其他局中人的行动;

(4) 不能确定其他局中人的推理。

为了对不确定情形下的决策建模,几乎所有的博弈论都使用了 von Neumann 和 Morgenstern[2] 以及 Savage[3] 的理论。也就是说,如果结果函数是随机的并被局中人已知(即,对每一个 $a \in A$,结果 $g(a)$是集合 C 上的一个不确定事件)(概率分布),那么局中人就被认为是为最大化一个函数期望值(v-N-M 效用)去行动,这个函数给每个结果赋一个值。如果行动与结果间的随机联系为给定,那么局中人就被认为是好像按其心中的一个概率分布去行动,这个分布决定了任何行动的结果。在这种情形下,局中人被认为将这样行动:即其心中有一个"状态空间"Ω,一个 Ω 上的概率测度,一个函数 $g: A \times \Omega \rightarrow C$ 和一个效用函数 $u: C \rightarrow R$;他被假定为考虑到概率测度去选择一个行动来最大化期望值 $u(g(a, w))$。

2.2.3 有限理性

在现实生活中谈论博弈时,通常关注各局中人能力的非对称性。例如,某些局中人可能对状况有更清楚的洞察力或有更强的分析能力。这些现实生活中如此关键的差异却被博弈论现有的形式忽略了。以象棋游戏为例:在一场真实的象棋游戏中,局中人在他们有关符合规则的行动知识方面和在分析能力方面都可能有差异。然而,如果用现代博弈论对象棋游戏建模,则会认为局中人具有的象棋规则知识是充分的,他们的分析能力也是理想的。对于"理性"局中人来说,象棋是一种平庸博弈,存在一种算法来"解"此博弈。该技术为两个局中人各确定了一个策略,这样就有一个"均衡"的结果,该结果的性质是无论其他局中人采用何种策略,只要局中人遵从他自己的策略,那么其结果至少与均衡结果一样好。这个结果(最早由 Zermelo[4] 发现)表明象棋游戏是没有什么意思的,因为它仅有一个可能结果。然而尽管有这一结论,象棋仍是一种有趣的游戏。它的均衡结果仍有待于计算,现在尚不可能采用 Zermelo 算法。即使有朝一日可以证明白方有一种取胜策略,而要一个人去实行它也是行不通的。由此抽象的象棋模型可推导出有关该博弈的一个重要事实,但同时又忽略了一场真实象棋游戏结果的最终决定因素——局中人的"能力"。

对不同局中人的能力及形势洞察力的不对称性建模在将来的研究中将是一个吸引人的挑战,对此,"有限理性"模型便开始发挥作用。

2.2.4　常见博弈模型

1. 囚徒困境博弈(prisoner's dilemma game，简称 PDG)

假设有两个嫌犯合伙作案之后被捕，但是警察并没有足够的证据指控两人有罪。因此，警察将两人分开审问，并提供了以下选择：如果两人都不检举对方(合作，cooperate，C)，则两人都会因为证据不足而得到轻判，获刑 3 年；如果两人都检举对方(背叛，defect，D)，则两人都会被判刑 5 年；如果一人检举对方，而另一人不检举，那么检举的一方会被判 1 年，而不检举的一方则会被重判获刑 10 年。其收益矩阵可见表 2-2。

表 2-2　囚徒困境博弈的收益矩阵

乙 \ 甲	合作	背叛
合作	(3,3)	(1,10)
背叛	(10,1)	(5,5)

2. 雪堆博弈(snowdrift game，简称 SDG)

雪堆博弈是另一类两人对称博弈模型，描述了两个人相遇时是彼此合作共同受益，还是彼此欺骗来相互报复。它揭示了个体理性和群体理性的矛盾对立。

假设在一个风雪交加的夜晚两人开车相向而行，被一个雪堆所阻挡。假设铲雪使道路畅通所付出的成本是 4，而道路畅通带来的好处是 10。当两人都选择铲雪(合作)时，他们的收益是 8；当一人选择铲雪，另一人待在车中等待(背叛)时，合作者的收益是 6，而背叛者的收益是 10；当两人都待在车中等待时，他们都无法回家，获得收益是 0。其收益矩阵见表 2-3。

表 2-3　雪堆博弈的收益矩阵

乙 \ 甲	合作	背叛
合作	(8,8)	(6,10)
背叛	(10,6)	(0,0)

雪堆博弈是一种对称博弈，个体的最优选择与对手的策略有关，即个体选择与对手不同的策略才能使自己的收益最大化。雪堆模型与囚徒困境不同的是，遇到背叛者时合作者的收益高于双方相互背叛的收益。因此，一个人的最佳策略取决于对手的策略：如果对手选择合作，他的最佳策略是背叛；反过来，如果对手选择背叛，那么他的最佳策略是合作。这样合作在系统中不会消亡，与囚徒困境相比，合作更容易在雪堆博弈中涌现。

3. 协调博弈(coordination game)

协调博弈的最经典案例出现在“交通博弈”中。两辆车甲和乙相向而行，在某个地点相

遇。如果两辆车各走各的左边或右边,那么两辆车都能顺利通过,则两辆车各获得 1 个收益;如果甲车走自己的右边,而乙车走自己的左边,那么两辆车就会相撞,则两辆车各获得 0 个收益。收益矩阵如表 2-4 所示。如果两辆车都希望顺利通过,不愿发生相撞甚至交通事故,那么,是靠右走还是靠左走呢?如果甲车知道乙车的选择,那么他就很容易做出自己的选择,反过来,乙车也想知道对方的选择。协调博弈强调局中人行为选择信念的一致性。

表 2-4　协调博弈的收益矩阵

乙＼甲	靠左	靠右
靠左	(1,1)	(0,0)
靠右	(0,0)	(1,1)

4. 公共品博弈(public goods game,简称 PGG)

公共品博弈是一种多人博弈。假设有 N 个局中人参与博弈,每个局中人都拥有一定数额的初始资金,可以选择投入任意额度的资金到公共资金中(合作),也可以不投入(背叛)。经过运作之后公共资金增值一定倍数,然后将总的公共资金平均分配给所有参与博弈的局中人。局中人投入资金越大,公共资金积累越多,局中人的收益相应就越多。如果所有局中人都选择投入自己的全部资金,则整体利益获得最大。但是每个局中人都面临着搭便车(free rider)的诱惑,即不投入也能够获得收益。因此,理性的局中人会选择不投入资金;而如果所有局中人都不投入资金,那么所有局中人都没有任何收益。

假设 3 人参加一个项目,每个人都有初始资金 100 元。其中,甲很小气,还想搭个便车,于是他投资 0 元;乙对项目的信心不足,于是他投资 60 元;而丙则对项目信心十足,他投资了全部资金 100 元。那么公共账户就有 160 元,接下来 160 元增值成为 330 元,再平均分成 3 份,甲、乙、丙每人都得 110 元。最后甲有 210 元,乙有 150 元,丙有 110 元。可以看出丙投入最大获益最小,而甲投入最小获益最大,搭便车才是局中人利益最大化的选项。

5. 鹰鸽博弈(hawk dove game)

鹰鸽博弈是指进行对抗时,两种不同的对抗策略,或如鹰般凶猛强硬,或如鸽子般温和隐忍。鹰搏斗时凶狠霸道,除非身负重伤,否则绝不退缩。而鸽的惯常方式威胁恐吓,从不伤害对手,往往委曲求全。因此,当两只鹰同时发现食物进行搏斗时,就会一直打到其中一只受重伤或死亡才罢休,它们的收益分别是−2;当鹰和鸽同时发现食物时,鹰要战斗,而鸽放弃食物逃跑,鸽不会受伤,这样鹰的收益是 2,而鸽的收益是 0;当两只鸽同时发现了食物时,它们会友好地分享食物,收益分别是 1。收益矩阵如表 2-5 所示。

表 2-5　鹰鸽博弈收益矩阵

乙＼甲	鹰策略	鸽策略
鹰策略	(−2,−2)	(2,0)
鸽策略	(0,2)	(1,1)

鹰鸽博弈是研究同一物种、种群内部竞争和冲突中的策略与均衡问题,广泛应用于人类社会中普遍存在的竞争和冲突等现象。它是由 Maynard Smith 根据两种动物的习性所提出的,用于描述演化博弈。演化博弈相关内容将在第 3 章加以介绍。

2.3 纳什均衡

2.3.1 纳什均衡及举例

在博弈中使用最广泛的解的概念是纳什均衡。这个概念体现了策略博弈行动的稳定状态,在此状态下每一个局中人都拥有对其他局中人行动的正确预期,并能够理性行动。它并不试图去检查稳定状态达到的过程。

定义 2.4 策略博弈$\langle N,(A_i),(\phi_i)\rangle$的纳什均衡(a Nash equilibrium of a strategic game)是一个行动组合$\alpha^* \in A$,α^*的性质是:对每一个局中人$i\in N$,有

$$(\alpha_{-i}^*,\alpha_i^*)\geqslant(\alpha_{-i}^*,\alpha_i),\quad 对所有\ \alpha_i\in A_i$$

这样,对α_i^*为一纳什均衡而言它必须满足:对其中任何一个局中人i,当其他每个局中人j选择均衡行动$\alpha_j{}^*$时,局中人i没有其他行动产生的结果优于其选择α_i^*所产生的结果。简言之,若给定其他局中人的行动,则局中人没有积极性选择其他行动。纳什证明了有限次的标准形式的博弈至少存在一个纳什均衡。但是,Daskalakis 等人证明了纳什均衡是 NP 问题的子集,不是通常认为的 NP-完全问题,而是 PPAD-完全问题(polynomial parity argument for directed graphs,简称 PPAD)。

下面对定义的重新表述有时是有用的。对任意$a_{-i}\in A_{-i}$,定义$B_i(a_{-i})$为局中人i在给定a_{-i}下的最佳行动集合:

$$B_i(a_{-i})=\{a_i\in A_i:(a_{-i},a_i)\geqslant i(a_{-i},a')\},\quad 对所有a'\in A_i$$

集值函数B_i称为局中人i的最优反应函数。纳什均衡是一个行动组合a^i,满足:

$$a_i^*\in B_i(a_{-i}^*),\quad 对所有\ i\in N$$

这个定义的代替形式提供了(不一定有效)一种寻找纳什均衡的方法:首先计算每个局中人的最优反应函数,然后再寻找一个行动组合a^*满足:对全部$i\in N$,有$a_i^*\in B_i(a_{-i}^*)$。若函数B_i是单值的,则第 2 步就可解具有$|N|$个未知数$(a_i^*)_{i\in N}$的$|N|$个方程。

下面的经典博弈代表了一些策略情形。这些博弈非常简单,在每个博弈里仅有两个局中人并且每个局中人仅有两个可能行动。但是,每个博弈都具有一类策略相互作用的本质特征,这些本质特征在更复杂的情形下也是经常可见的。

例 2.1 囚徒困境。

一个案子的两个嫌疑犯被分开审讯,警官分别告诉两个囚犯,如果你招供,而对方不招供,则你将被判刑 1 年,而对方将被判刑 10 年;如果两人均招供,将均被判刑 5 年。于是,两人同时陷入招供还是不招供的两难处境。其实,如果两人均不招供,将最有利,只被判刑 3 个月。于是得到的博弈矩阵如表 2-6 所示。

表 2-6　囚徒困境的博弈矩阵

囚犯乙＼囚犯甲	招供	不招供
招供	判刑 5 年	甲判刑 10 年；乙判刑 1 年
不招供	甲判刑 1 年；乙判刑 10 年	判刑 3 个月

这是一种通过合作可获利的博弈——对局中人来说最好的结果是两个人都不招供——每个局中人都成为无罪释放者。但无论其中一人怎么做，另一个人都偏好于招供而不是不招供，这时博弈有唯一纳什均衡（招，招）。囚徒困境反映了个体的最佳选择并不是群体的最佳选择，也就是说，个体最佳利益与群体最佳利益存在不一致性。

例 2.2　军备竞赛。

美苏两国是世界上公认的两个超级大国，两个国家都有两个策略供选择，一个是扩军备战，发展核武器；另一个是彻底裁军，和平共处。扩军每年需要 2000 亿美元用于军备，彻底裁军则军费为 0。但如果一方裁军，另一方扩军，裁军的一方则冒有国家被颠覆的风险，得益为$-\infty$，而扩军的一方则可能从被颠覆的一方掠夺到有限数量的财富，假定为 10 000 亿美元，扣除军费开支，得益为 8000 亿美元。于是得到的博弈矩阵如表 2-7 所示。

表 2-7　军备竞赛的博弈矩阵

美国＼苏联	扩军	裁军
扩军	$-2000,-2000$	$8000,-\infty$
裁军	$-\infty,8000$	0,0

采用与上述相同的分析方法，很显然（扩军，扩军）是该博弈的纳什均衡。因为，对于每一个博弈方来说，无论对方采取何种策略，自己采用扩军策略总是有利的。显然，组合（裁军，裁军）与均衡（扩军，扩军）相比无疑要好得多，但是，由于双方都从各自利益出发，因而最终得到了从全局看最不利的均衡。

例 2.3　智猪博弈。

猪圈里有一头大猪和一头小猪，猪圈的一头有一个饲料供应的按钮，可由大猪和小猪按下，猪槽在猪圈的另一头。按一下按钮就会有 10 个单位饲料进槽，但谁按就要付出 2 个单位的成本，比如体力的消耗。若大猪先到，则大猪吃到 9 个单位，小猪吃到 1 个单位；若同时到，则大猪吃 7 个单位，小猪吃 3 个单位；若小猪先到，则大猪吃 6 个单位，小猪吃 4 个单位。各种情景组合扣除成本后的赢得矩阵可如表 2-8 所示。

表 2-8　智猪博弈的博弈矩阵

大猪＼小猪	按	等待
按	5,1	4,4
等待	$9,-1$	0,0

从这个例子可以发现,大猪选择按,小猪最好选择等待;大猪选择不按,小猪还是最好选择等待。即无论大猪选择按还是不按,小猪的最佳策略都是等待。也就是说,无论如何,小猪都只会选择等待。在这样的情况下,大猪最好的选择是按,因为不按的话都饿肚子,按的话还可以有 4 个单位的收益。所以纳什均衡是(大猪按,小猪等待)。

例 2.4 斗鸡博弈。

有两个格斗士约定,两人各驾一辆小车在某个公众场合相向对开,谁退让谁就输,输者支付赢者 10 万美元。这种做法着实有些冒险,如果两人互不相让,则结果一定是同归于尽,双方的收益各为负无穷大;如果双方都退让,则各自收益为 0。其博弈矩阵见表 2-9。

表 2-9 斗鸡博弈的博弈矩阵

格斗士甲 \ 格斗士乙	冲撞	退让
冲撞	$-\infty,-\infty$	$10,-10$
退让	$-10,10$	$0,0$

用划线法或箭头法容易得到,该博弈有两个纳什均衡,即(冲撞,退让)或(退让,冲撞)。

2.3.2 纳什均衡的存在性

就像猜谜博弈那样,不是每个策略博弈都有纳什均衡。博弈的纳什均衡集合非空的条件已被广泛考察过,下面给出此类结论中最简单的一种。

存在性结论有两个目的。首先,如果有一个符合结论假设的博弈,那么为寻求均衡付出的努力就会有成功的希望。其次,更重要的是,均衡存在性表明博弈是与一个稳定状态解相一致的。进一步地,一类博弈的均衡存在性使得研究者可以去研究均衡的性质而不用具体地找到均衡,也不用冒研究空集的风险。

为了说明一个博弈具有纳什均衡就不得不去说明行动组合 a^*,其满足对所有 $i\in N$,$a_i^*\in B_i(a_{-i}^*)$。通过 $B(a)=X_{i\in N}B_i(a_{-i})$来定义集值函数 B: $A\rightarrow A$。那么可以简单地写成向量形式 $a^*\in B(a^*)$。不动点定理给出了 B 上的条件,在此条件下确定存在一个值 a^* 满足 $a^*\in B(a^*)$。

引理 2.1(Kakutani 的不动点定理[5]) 令 X 是$\mathbb{R}^n$的一个紧凸子集,令 f:$X\rightarrow X$ 是一个集值函数且满足:

- 对所有 $x\in X$,集合 $f(x)$非空且凸;
- f 的图形是闭的(亦即:对所有序列$\{x_n\}$和$\{y_n\}$,对所有 n 有 $y_n\in f(x_n)$,$x_n\rightarrow x$,$y_n\rightarrow y$,有 $y\in f(x)$)。

那么,存在 $x^*\in X$ 满足 $x^*\in f(x^*)$。

命题 2.1 策略博弈$\langle N,(A_i)(\succsim_i)\rangle$若对所有 $i\in N$ 满足:

局中人 i 的行动集合 A_i是欧氏空间的一个非空紧凸子集;并且偏好关系$\succsim_i$是

(1) 连续的;

(2) 在 A_i上拟凹的,

则该博弈有一个纳什均衡。

证明：通过 $B(a)=X_{i\in N}B_i(a_{-i})$ 来定义 $B: A\to A$（这里，B_i 是局中人 i 的最优反应函数）。对任一 $i\in N$，由于 $\succsim_i$ 是连续的且是紧的，则 $B_i(a_{-i})$ 非空，又由于 $\succsim_i$ 在 A_i 是拟凹的，所以 $B_i(a_{-i})$ 又是凸的；因为每个 $\succsim_i$ 是连续的，所以 B 有闭图。这样，应用 Kakutani 定理 B 有一个不动点，任何不动点都是博弈的纳什均衡。□

注意：这个结论保证了满足一定条件下的策略博弈至少有一个纳什均衡，一个博弈可以有不止一个均衡。同时也应注意，该命题不适用于任何其中有某个局中人有有限多个行动的博弈，因为这种博弈违背了每个局中人行动集合是凸的条件。

2.3.3 严格竞争博弈

对于一个任意的策略博弈纳什均衡，仅在一些确定的博弈类型里能够对均衡的特性做些说明。一种确定类型的博弈是这样的，在这种博弈中有两个局中人，他们的偏好是正好相反的。为讨论方便，下面假定局中人的名字为“1”和“2”。

定义 2.5　策略博弈 $\langle\{1,2\},(A_i),(\succsim_i)\rangle$ 是严格竞争的(strictly competitive)，如果对任一 $a\in A$ 和 $b\in A$，有 $a\succsim_1 b$，当且仅当 $b\succsim_2 a$。

一个严格竞争博弈有时也叫零和(zero sum)博弈，因为用支付函数 u_1 表示局中人 1 的偏好关系 $\succsim_1$，用 u_2 表示局中人 2 的偏好关系，则有 $u_1+u_2=0$。

对于局中人 i，如果他在假设无论自己采取什么行动，局中人 j 都将选择使他损失尽可能大的行动的基础上选择对他自己最好的行动，则局中人 i 的行动是最大最小化行动(max minimizes)。需要说明的是，对于一个具有纳什均衡的严格竞争博弈来说，一对行动是纳什均衡当且仅当每个局中人的行动都是最大最小化行动。这个结论很有意义，因为它提供了一种个人决策与纳什均衡概念背后的理论之间的联系。在推导这一结论的过程中也要证明一个具有纳什均衡的严格竞争博弈的所有均衡产生同一支付这一强结论。在非严格竞争博弈中纳什均衡的这个特性是很少能得到满足的。

定义 2.6　令 $\langle\{1,2\},(A_i),(u_i)\rangle$ 是一个严格竞争策略博弈。行动 $x^*\in A_1$ 是局中人 1 的最大最小化行动，如果有

$$\min_{x\in A_2} u_1(x^*,y)\geqslant \min_{x\in A_2} u_1(x,y),\quad \text{对所有 } x\in A_1$$

同样，行动 $y^*\in A_2$ 是局中人 2 的最大最小化行动，如果有

$$\min_{x\in A_1} u_2(x,y^*)\geqslant \min_{x\in A_1} u_2(x,y),\quad \text{对所有 } y\in A_2$$

简言之，对局中人 i 来说，最大最小化行动就是最大最小化局中人 i 能够保证获得的支付的行动。对局中人 i 来说，最大最小化行动解决了问题 $\max_x \min_y u_1(x,y)$，相应地，对局中人 2 来说，最大最小化行动解决了问题 $\max_y \min_x u_2(x,y)$。

在下文里，为了方便起见，假定局中人 1 的偏好关系用支付函数 u_1 表示，并且不失一般性，$u_2=-u_1$。下面的结论表明，局中人 2 支付的最大最小化值等同于局中人 1 支付的最小最大化值。

引理 2.2　令 $\langle\{1,2\},(A_i),(u_i)\rangle$ 为一个严格竞争策略博弈。那么有

$$\max_{y\in A_2}\min_{x\in A_1} u_2(x,y)=-\min_{y\in A_2}\max_{x\in A_1} u_1(x,y)$$

进一步而言，$y\in A_2$解决了问题$\max\limits_{y\in A_2}\min\limits_{x\in A_1}u_2(x,y)$当且仅当它解决了问题$\min\limits_{y\in A_2}\max\limits_{x\in A_1}u_1(x,y)$。

证明：对任何函数f，有$\min\limits_z(-f(z))=\max\limits_z f(z)$和$\mathrm{argmin}_z(-f(z))=\mathrm{argmax}_z(f(z))$。从而对每一 $y\in A_2$，有 $-\min\limits_{x\in A_1}u_2(x,y)=\max\limits_{x\in A_1}(-u_2(x,y))=\max\limits_{x\in A_1}u_1(x,y)$。这样，$\max\limits_{y\in A_2}\min\limits_{x\in A_1}u_2(x,y)=-\min\limits_{y\in A_2}[\min\limits_{x\in A_1}u_2(x,y)]=-\min\limits_{y\in A_2}\max\limits_{x\in A_1}u_1(x,y)$；并且有：$y\in A_2$ 是问题$\max\limits_{y\in A_2}\min\limits_{x\in A_1}u_2(x,y)$的解，当且仅当它是问题$\min\limits_{y\in A_2}\max\limits_{x\in A_1}u_1(x,y)$的解。 □

2.3.4 混合策略纳什均衡

混合策略纳什均衡概念被设计用来对某种博弈的稳定状态建模，在此种博弈中局中人的选择并不确定而是受概率规则调节。

一个策略博弈定义为一个三元组$\langle N,(A_i,(\succsim_i))\rangle$，这里，每个局中人 i 的偏好关系$\succsim_i$被定义在行动组合集合$A=X_{i\in N}A_i$上。在本节中，容许局中人的选择是非确定的，这样就需要给模型的基本元素增加一个每个局中人关于在 A 上不确定事件的偏好关系的详细说明。遵从于现代博弈理论的习惯，假定每个局中人 i 的偏好关系满足 von Neumann 和 Morgenstern 的假设，这样，偏好关系就可用某一函数 u_i：$A\to\mathbb{R}$ 的期望值来表示。因此，本节中关于策略相互作用的模型是一个三元组$\langle N,(A_i),(u_i)\rangle$，它不同于原来定义的策略博弈，这里对每一$i\in N$，$u_i$：$A\to\mathbb{R}$ 是一个期望值，代表了局中人 i 关于A 上的不确定事件集合的偏好关系的函数。不过，我们将这里给出的模型简称为策略博弈。

令$G=\langle N,(A_i),(u_i)\rangle$是一个这样的策略博弈：用 $\Delta(A_i)$来表示 A_i上的概率分布并且称 $\Delta(A_i)$的一个元素为局中人的一个混合战略(mixed strategy)。假定局中人 i 的混合战略是独立随机化的。为明确起见，也称 A_i的一个元素为一个纯策略(pure strategy)。对任一有限集 X 和$\delta\in\Delta(X)$，用$\delta(x)$表示 δ 赋给 $x\in X$ 的概率并且将δ 的 δ 支集定义为由 $x\in X$ 构成的集合，其中，$\delta(x)>0$。混合策略的一个 X $(\alpha_i)_{j\in N}$产生了集合A 上的一个概率分布。

现在从 G 派生出另一个策略博弈，称为 G 的“混合扩展”，在该博弈里每个局中人的行动集合是他在 G 中混合策略集合 $\Delta(A_i)$。

定义 2.7 策略博弈$\langle N,(A_i),(u_i)\rangle$的混合扩展(mixed extension)是这样一个策略博弈$\langle N,(\Delta(A_i)),(U_i)\rangle$；$\Delta(A_i)$是 A_i上的概率分布集合，并且U_i：$X_{j\in N}\Delta(A_j)\to\mathbb{R}$ 将 U_i下关于A 上由 α 引起的不确定事件的期望值赋给每个 $\alpha\in X_{j\in N}\Delta(A_j)$(这样，若 A 有限，则 $U_i(\alpha)=\sum\limits_{a\in A}(\prod\limits_{j\in N}\alpha_j(a_j))U_i(a))$。

请注意每个函数 U_i都是多重线性的。也就是说，对任意混合策略组合 α，局中人 i 的任何混合策略 β_i 和 Y_i 以及任一 $\lambda\in[0,1]$，有$U_i(\alpha_{-i},\lambda\beta_i+(1-\lambda)Y_i)=\lambda U_i(\alpha_{-i},\beta_i)+(1-\lambda)U_i(\alpha_{-i},Y_i)$。同时，要注意，在每一 A_i为有限情形下，对任一混合策略组合有

$$U_i(\alpha)=\sum_{a\in A}\alpha_i(a_i)U_i(\alpha_{-j},e(\alpha_i))$$

这里，$e(\alpha_i)$是局中人 i 的退化混合策略，它使得 $\alpha_i\in A_i$的概率为 1。

下面来定义本节要研究的主要均衡的概念。

定义 2.8　一个策略博弈的混合策略纳什均衡[6](mixed strategy Nash equilibrium)是该博弈混合扩展的纳什均衡。

假定$\alpha^* \in X_{j\in N}\Delta(A_j)$是$G=\langle N,(A_i),(u_i)\rangle$的一个混合策略纳什均衡,该均衡中每个局中人 i 的混合策略α_i在它将概率 1 赋给一个元素的意义下是退化的。那么,因为A_i可用$\Delta(A_i)$的一个子集来确定,所以行动组合α_i是G的一个纳什均衡。反之,假定α_i是G的一个纳什均衡,那么由u_i在α_i中的线性可知,没有A_i中的行动之上的概率分布使局中人i获得的支付高于由$e(\alpha_i^*)$产生的支付,于是组合$e(\alpha_i^*)$是G的混合策略。

一个策略博弈的纳什均衡集是它的混合策略纳什均衡集合的一个子集。存在纳什均衡集为空的博弈。对于局中人只有有限多个行动的博弈至少有一个混合策略纳什均衡。

命题 2.2　任何有限策略博弈都有一个混合策略纳什均衡。

证明:令$G=\langle N,(A_i),(u_i)\rangle$是一个策略博弈,并且对每一个局中人 i 令 m_i 为集合 A_i 元素的个数。使用向量$(p_1,\cdots,p_k,\cdots,p_m)$集合来确定局中人的混合战略集合$\Delta(A_i)$,其中,对所有$k$,$p_k\geqslant 0$且$\sum_{k=1}^{m_i} p_k=1$($p_k$为局中人$i$使用第$k$个纯策略的概率)。这个向量集合是非空、凸和紧的。既然在概率上期望支付是线性的,那么每个局中人在G的混合扩展中的支付函数对于他自己的策略是拟凹和连续的。这样,G的混合扩展满足命题所有的条件。□

下面的结论给出了纳什均衡的一个非常重要的性质,它在计算均衡时很有用。

引理 2.3　令$G=\langle N,(A_i),(u_i)\rangle$为一个有限策略博弈。那么,$\alpha^*\in X_{i\in N}\Delta(A_i)$为$G$的一个混合策略纳什均衡,当且仅当对每个局中人$i\in N$,$\alpha_i^*$的子集上的每个纯策略是对$\alpha_{-i}^*$的最优反应。

证明:先假定在α_i^*子集中有一个行动α_i,它不是α_{-i}^*的最优反应。接着,由U_i在α_i中的线性性质,局中人i可通过从α_i向一个最优反应行动转移概率来增加其支付;因此,α_i^*不是对α_{-i}^*的最优反应。

然后假定有一混合策略α_i'提供的期望支付高于α_i^*对α_{-i}^*的反应支付。由于U_i是线性的,因此至少有一个α_i'子集上的行动一定能比α_i^*子集上的行动提供更高的支付,因此并不是α_i^*支集上所有行动都是α_{-i}^*的最优反应。

从而有,在任意一个局中人的均衡混合策略子集中的每个行动都使该局中人产生同样的支付。□

例 2.5　委托代理博弈。

该模型也称为监察博弈模型,描述委托人应该如何处理与代理人之间的关系。代理人为委托人干活,可以有两个策略供选择,或者偷懒或者努力工作。工作需要支付一定的代价或费用成本g,同时获得委托人支付的工资$w(w>g)$。委托人在监督监察方面也有两个选择,即检查或不检查。检查需要支付一定的费用h(假定$0<h<g$),但坚持检查在代理人努力工作的条件下可以使公司不断增值,假如增加到$v(v>w)$,同时在检查中如发现代理人偷懒,则以扣除工资作为惩罚,于是得到的博弈矩阵如表 2-10 所示。

表 2-10　委托代理的博弈矩阵

代理人＼委托人	检查	不检查
偷懒	$0, -h$	$w, -w$
工作	$w-g, v-w-h$	$w-g, v-w$

根据假定的代数关系，不难判断该博弈没有纯策略意义下的纳什均衡。先假定代理人偷懒的概率为 x，则工作的概率为 $1-x$；假定委托人检查的概率为 y，则不检查的概率为 $1-y$。于是，x 的选择就应该满足

$$-hx+(v-w-h)(1-x)=-wx+(v-w)(1-x)$$

解之，得到：$x=h/w$。

而 y 的选择则应该满足

$$w(1-y)=(w-g)y+(w-g)(1-y)$$

解之，得到：$y=g/w$。

分析结果提供了非常重要的参考信息：作为委托人，要降低代理人偷懒的概率，首先必须增加支付代理人的工资 w，以充分调动其工作的积极性；其次，要降低检查费用，也就是要少检查，即相信代理人，尊重代理人。而要降低检查的概率，除了首先增加代理人工资以外，还必须设法降低代理人为获取工资 w 而支出的费用成本，这就要求尽可能地改善代理人的福利待遇，努力搞好各方面必要的服务等等。

2.4　合作博弈

2.4.1　可转移支付合作博弈

从一个简单的联盟博弈形式开始，其中每个局中人群都与唯一的一个数相联系，该数被解释为对人群有效的支付；对于这个支付如何在成员间进行分配没有任何限制。

定义 2.9　一个可转移支付联盟博弈(coalitional game with transferable payoff)包括：

① 一个有限集合 N（局中人集合）；

② 将 N 的每个非空子集 S（一个联盟）与某个实数 $v(S)$（S 的值(worth)）相联系的一个函数 v。

对每个联盟 S，实数 $v(S)$是在 S 成员间有效分配的全部支付。即，联盟 S 能够采取的联合行动集合包括在 S 成员间所有可能的 $v(S)$分配。

在本节，假定研究的可转移支付联盟满足下列条件：所有局中人的联盟 N 的值至少与 N 的任一分割的全部元素值的总和一样多。这个假设保证了所有局中人的联盟 N 的形成是最优的。

定义 2.10　一个可转移支付联盟博弈〈N，v〉是凝聚性的，如果对 N 的每个分割 $\{S_1, S_2, \cdots, S_k\}$，有 $v(N)\geqslant\sum_{k=1}^{k}v(S_k)$。

这是超可加性条件的一个特例，它要求满足 $S \cap T = \varnothing$ 的联盟 S 和 T 有 $v(S \cup T) \geqslant v(S) + v(T)$。

2.4.2　核

核的概念所蕴含的思想类似于非联盟博弈的纳什均衡所蕴含的思想：如果没有偏离是有利的，则这个结果是稳定的。在核的情形中，一个结果是稳定的，如果没有联盟能够偏离并获得一个对其所有成员都更好的结果。对于一个可转移支付联盟博弈来说，保持稳定性的条件是联盟不能获得一个超过其成员现有支付之和的支付。给定关于博弈是凝聚性的假设，将限于这样的结果，即所有局中人的联盟 N 形成。

令 $\langle N, v\rangle$ 为一个可转移支付联盟博弈。对任一实数组合 $(x_i)_{i \in N}$ 和任一联盟 S，令 $x(S) = \sum_{i \in S} x_i$。如果 $x(S) = v(S)$ 是一个实数向量，$(x_i)_{i \in S}$ 是一个 S-可行支付向量（S-feasible payoff vector），则称一个 N-可行的支付向量为一可行支付组合（feasible payoff profile）。

定义 2.11　可转移支付联盟博弈 $\langle N, v\rangle$ 的核是可行支付组合 $(x_i)_{i \in S}$ 的集合，对于它没有满足对所有 $i \in S$ 有 $y_i > x_i$ 的联盟 S 和 S-可行支付向量 $(y_i)_{i \in S}$。

一个很明显的等价定义是：核是一个满足对每个联盟 S 有 $v(S) \leqslant x(S)$ 的可行支付组合 $(x_i)_{i \in S}$ 的集合。所以，核是一个满足弱线性不等式系统的支付组合集合，因此是闭的和凸的。

下面给出的例子表明了范围较广的可被模块化为联盟博弈范围的情形，并且阐述了核的概念。

例 2.6　3 人多数博弈（three-player majority game）。假定 3 个局中人可获得 1 单位支付，他们中的任何两个若有独立于第 3 人的行动可获得 $a \in [0,1]$，任何单独一个独立于其他两个人的行动将一无所得。可将这个情形模化为联盟博弈，其中，$N = \{1,2,3\}$，$v(N) = 1$，只要 $|S| = 2$，则 $v(S) = a$，并且对所有 $i \in N$ 有 $v(|i|) = 0$。这个博弈的核是所有非负支付组合 (x_1, x_2, x_3) 的集合，对于它，有 $x(N) = 1$ 及对每个局中人的联盟 S 有 $v(S) \geqslant a$。因此核的非空当且仅当 $a \leqslant \frac{2}{3}$。

例 2.7　一支探险队发现了山中的宝藏，他们中的每一对都能拿到一份。模仿这个情形的联盟博弈是 $\langle N, v\rangle$，这里，

$$v(S) = \begin{cases} \dfrac{|S|}{2}, & \text{若 } |S| \text{ 是偶数} \\ \dfrac{|S| - 1}{2}, & \text{若 } |S| \text{ 是奇数} \end{cases}$$

如果 $|N| \geqslant 4$ 是偶数，那么核包括唯一的支付组合 $\left(\frac{1}{2}, \cdots, \frac{1}{2}\right)$。如果 $|N| \geqslant 3$ 是奇数，那么核是空的。

2.4.3　核的非空性

现在导出一个联盟博弈核非空的条件。因为核是由一个线性不等式系统定义的，所以这样一个条件可从一般不等式系统解的存在条件导出。不过，因为用于定义核的不等式系统有一个特别的结构，所以能导出一个更具体的条件。

用 C 表示所有联盟集合，对任一联盟 S 用 $\mathbb{R}^S$ 表示维数由 S 的成员确定的 $|S|$ 维欧氏空间，给定

$$(1_S)_i = \begin{cases} 1, & \text{若 } i \in S \\ 0, & \text{否则} \end{cases}$$

用 $1_S \in \mathbb{R}^N$ 表示 S 的特征向量。$[0,1]$的数族$(\lambda_S)_{S\in C}$是一个平衡权数族，如果对每个局中人 i 关于所有包含的联盟 λ_S的和是 1：$\sum_{S\in C}\lambda_S 1_S = 1_N$。作为一个例子，令 $|N|=3$。那么：若 $|S|=2$，则$\lambda_S=\frac{1}{2}$，否则，$\lambda_S=0$ 的数族(λ_S)是一个平衡权数族；满足 $|S|=1$ 则$\lambda_S=1$，否则 $\lambda_S=0$ 的数族(λ_S)也是平衡权数族。博弈$\langle N,v\rangle$是平衡的(balanced)，如果对每个平衡权数族有 $\sum_{S\in C}\lambda_S v(S)\leqslant v(N)$。

关于平衡博弈概念的一个解释如下，每个局中人有 1 单位时间，他必须在所有其所在的联盟间将它进行分配。为了使在 λ_S部分时间内联盟 S 是活动的，则它的所有成员在这个时间部分也必须是在 S 中活动的，在此情形中，联盟产生支付 $\lambda_S v(S)$。在这个解释中，权数族为平衡的条件是一个关于局中人时间分配的可行条件。并且，一个博弈是平衡的，如果没有给局中人产生多于 $v(N)$支付的可行时间分配。

下面的结论被称为 Bondareva-shapley 定理[7,8]。

命题 2.3　一个可转移支付联盟博弈有一个非空核当且仅当它是平衡的。

证明：令$\langle N,v\rangle$为一个可转移支付联盟博弈。首先，令 x 为$\langle N,v\rangle$核中一个支付组合，并且令$(\lambda_S)_{S\in C}$为一个平衡权数族。那么，$\sum_{S\in C}\lambda_S v(S)\leqslant \sum_{S\in C}\lambda_S x(S) = \sum_{i\in N}x_i\sum_{i\in S}\lambda_S = \sum_{i\in N}x_i = v(N)$，所以$\langle N,v\rangle$ 是平衡的。

现在假设$\langle N,v\rangle$是平衡的。那么没有平衡权数族$(\lambda_S)_{S\in C}$使得 $\sum_{S\in C}\lambda_S v(S) > v(N)$。因此，凸集$\{(1_N,v(N)+\varepsilon)\in \mathbb{R}^{|N|+1}: \varepsilon>0\}$不交于凸椎。

$\{y\in\mathbb{R}^{|N|+1}: y=\sum_{S\in C}\lambda_S(1,v(S))\}$，这里对所有 $S\in C$ 有$\lambda_S\geqslant 0\}$，因为若不是这样，那么$1_N=\sum_{S\in C}\lambda_S 1_S$，因此$(\lambda_S)_{S\in C}$ 是一个平衡权数族并且$\sum_{S\in C}\lambda_S v(S) > v(N)$。于是分离超平面定理有一个非零向量$(\alpha_N,\alpha)\in\mathbb{R}^N\times\mathbb{R}$，使得对所有椎体中的 y 和所有 $\varepsilon>0$ 有

$$(\alpha_N,\alpha)\cdot y \geqslant 0 > (\alpha_N,\alpha)\cdot(1_N,v(N)+\varepsilon)$$

因为$(1_N,v(N))$在核中，因此有 $\alpha<0$。

现在令 $x=\alpha_N/(-\alpha)$。因为$(1_S,v(S))$对所有 $S\in C$ 是在椎体中的，所以由左边的不等式有 $x(S)=x\cdot 1_S\geqslant v(S)$，且由右边的不等式有 $v(N)\geqslant 1_{Nx}=x(N)$。给 x 加上一个元素为非负的向量，得到 $v(N)=x(N)$，得到在$\langle N,v\rangle$的核中的一个支付组合。□

2.4.4　无可转移支付的合作博弈

在一个可转移支付联盟博弈中，每个联盟 S 是以一维的一个数 $v(S)$为显著特点的，其解释是：$v(S)$是一个可用任何方法在 S 的成员间进行分配的支付。现在研究一个更一般的概念，在其中每个联盟不必得到某个固定支付的所有分配；而且，每个联盟 S 是以一个任意的结果集合 $V(S)$为显著特征的。

定义 2.12　一个无可转移支付联盟博弈包括：

- 一个有限集合 N（局中人的集合）；
- 一个集合 X（结果集合）；
- 一个对 N 的每一非空子集 S（一个联盟）赋一个集合 $V(S)\subseteq X$ 的函数 V；
- 对每个局中人 $i\in N$ 有一个 X 上的偏好关系 $\succsim_i$。

任意一个可转移支付联盟 $\langle N,v\rangle$ 可如下与一个一般的联盟博弈 $\langle N,X,V,(\succsim_i)_{i\in N}\rangle$ 相联系：$X=\mathbb{R}^N$，对每个联盟 S 有 $V(S)=\left\{x\in\mathbb{R}^N:\sum_{i\in S}x_i=v(S)\right\}$，并且，若 $j\in N/S$，则有 $x_j=0$，$x\succsim_i y$ 当且仅当 $x_i\geqslant y_i$。在这种联系下，可转移支付联盟博弈集合是所有联盟集合的一个子集合。

一般联盟博弈的核定义是对可转移支付博弈的核的定义的自然推广。

定义 2.13　联盟博弈 $\langle N,V,X,(\succsim_i)_{i\in N}\rangle$ 的核是所有 $x\in V(N)$ 的集合；对于它，不存在联盟 S 和满足对所有 $i\in S$ 有 $y\succsim_i x$ 的 $y\in V(S)$。

在可转移支付联盟博弈的平衡条件下，一般联盟博弈的核是非空的。这里，我们不讨论这些条件。

2.4.5　Shapley 值

在一个联盟 τ 中，局中人 i 应获取的效用将取决于他对联盟 τ 的贡献度。定义 $m_i(\tau)=v(\tau)-v\left(\frac{\tau}{\{i\}}\right)$，$i\in\tau$，为局中人 i 对联盟 τ 的边际贡献，其中，$\frac{\tau}{\{i\}}$ 代表除 i 之外 τ 中其他人形成的联盟。用 τ 来衡量局中人 i 的加入能在多大程度上提高原联盟 $\frac{\tau}{\{i\}}$ 的总效用。这里，不能简单地将 $m_i(N)$ 作为 (N,v) 合作博弈下局中人 i 所得的效用分配，因为不能保证 $\sum_{i\in N}m_i(N)\leqslant v(N)$。*Shapley* 的构思则是：局中人 i 在 (N,v) 博弈下可能形成各种联盟，只要将他对不同联盟的边际贡献平均起来，就是其在全体联盟 N 下应得的最终效用分配 $\phi_i(N,v)$。称 $\phi_i(N,v)$ 为 (N,v) 博弈下的 Shapley 值，下面是它的严格定义。

定义 2.14　一个 (N,v) 合作博弈的 Shapley 值是一组分配 $\Phi=(\phi_1,\phi_2,\cdots,\phi_N)$，其中，$\phi_i(N,v)=\sum_{\tau\subset N}\frac{(|\tau|-1)!\,(N-|\tau|)!}{N!}m_i(\tau)$。

定义 2.14 中的 Shapley 值看起来复杂且不易理解。这个平均数为什么要这么构造？Shapley 任务合作局中人的分配应遵循以下 4 个公理。

一是效率公理，即 $\sum_{i\in N}\phi_i(N,v)=v(N)$，所有的效用都被分配无浪费。

二是虚拟公理，若局中人 i（虚拟局中人）对任何联盟均无贡献，即对 $\tau\subset N$，$m(\tau)=0$，那么其分配 $\phi_i(N,v)=0$。

三是对称公理，如果两个博弈 (N,v) 和 (N,v') 完全相同，只是局中人 i,j 的角色互换，比如资本所有者变成工人，工人变成资本所有者，那么，$\phi_i(N,v)=\phi_i(N,v')$，即标识并不改变分配。

四是加总公理，对于两个合作博弈 (N,v) 与 (N,v')，定义特征函数 $v+v'$ 为 $(v+v')(\tau)=v(\tau)+v'(\tau)$，其中，$\tau$ 是任意联盟。那么，对所有 $i\in N$，$\phi_i(N,v+v')=\phi_i(N,v)+\phi(N,v')$。

加总公理规定了不同合作博弈的分配 Φ 是如何相互联系的。人们需在这 4 个公理的前提下，寻找合适的分配方法。

Shapley 证明：存在唯一的 Shapley 值同时满足这 4 个公理，即以这 4 个公理为基础的合作博弈框架能够带来唯一的预测。那么，理论上合适的 Shapley 值在直观上又如何理解呢？下面以资本-劳动博弈为例进行说明。

考虑资本-劳动博弈中的局中人 1，他可能形成的联盟有{1}，{1，2}，{1，3}，{1，2，3}。每个联盟都可以计算出其边际贡献，局中人 1 对这 4 种联盟的边际贡献为 $m_1(\{1\})=v(\{1\})-v(\phi)=0$；$m_1(\{1,2\})=v(\{1,2\})-v(\{2\})=p$；$m_1(\{1,3\})=v(\{1,3\})-v(\{3\})=p$；$m_1(\{1,2,3\})=v(\{1,2,3\})-v(\{2,3\})=2p$。现在只要找出一个联盟发生的概率就能完成平均边际贡献值的计算。关键是这些概率如何确定？确定一个联盟发生的概率必须有一个合理的故事。可以想象各局中人以某种顺序进入一个房间，当局中人 1 进入后，总是和先于他进入的所有人结成联盟。局中人进入房间的顺序共有 3！＝6 种，即(1，2，3)，(1，3，2)，(2，1，3)，(2，3，1)，(3，1，2)和(3，2，1)，括号内的数依次代表最先、其次以及最后进入房间的人。假设每种进入顺序发生的可能性都相同。按此规则，局中人 1 能够结成联盟{1}的情形有第 1、第 2 两种顺序，其发生的概率为$\frac{2}{3!}$；结成联盟{1，2}的情形只有第 3 种顺序，发生概率为$\frac{1}{3!}$；结成联盟{1，3}的情形只有第 5 种顺序，发生概率为$\frac{1}{3!}$；而结成联盟(1，2，3)的情形只有第 4 和第 6 这两种顺序，发生概率为$\frac{2}{3!}$。以上穷尽了所有可能，概率加总正好为 1。于是，对局中人 1 来说，总的期望边际贡献 $\phi_1(v)=\frac{2}{3!}\times 0+\frac{1}{3!}\times p+\frac{2}{3!}\times 2p=p$，这就是局中人 1 的 Shapley 值。现在将这种平均边际贡献的思想一般化。对于一个(N,v)博弈来说，局中人 i 按以上规则形成联盟 τ 的概率是$\frac{(|\tau|-1)!(N-|\tau|)!}{N!}$，其中，$N!$ 是 N 人排列的种类，$(|\tau|-1)!$ 是出现在 i 之前的局中人排列的种类数，$(N-|\tau|)!$ 则是出现在 i 之后的局中人排列的种类数。这样，平均边际贡献就可以定义为 $\sum_{\tau\subset N}\frac{(|\tau|-1)!(N-|\tau|)!}{N!}m_i(\tau)$。下面给出一个计算 Shapley 值的例子。

例 2.8 资本-劳动博弈的 Shapley 值。

解：局中人 1 的 Shapley 值前文已求出，为 p。局中人 2，3 对称，其 Shapley 值相等。这里以局中人 2 为例，他能够结成的联盟也有 4 种，{2}，{1，2}，{2，3}，{1，2，3}。他对 4 种联盟的边际贡献分别为 $0,p,0,p$。根据定义 2.14，4 种联盟出现的概率分别为$\frac{2}{3!}$，$\frac{1}{3!}$，$\frac{1}{3!}$和$\frac{2}{3!}$，所以，局中人 2 的 Shapley 值为 $\phi_2(v)=\frac{2}{3!}\times 0+\frac{1}{3!}\times p+\frac{1}{3!}\times 0+\frac{2}{3!}\times p=p/2$。这样，资本-劳动博弈的 Shapley 值为 $\phi=(p,p/2,p/2)$。注意到，$v(n)=2p$，Shapley 值的预测即是资本独取一般效用，两人则平分另一半。当有 N 个工人时，每个工人的 Shapley 值是 p/N。

小结

博弈是研究人与人之间存在着利益冲突时，局中人所进行的行为选择；而博弈论正是运

用现代的数学模型来研究博弈行为的理论。

- (纳什均衡的存在性)所有这样的博弈都有一个纳什均衡,即每个局中人的行动集合是紧的和凸的,并且每个局中人的偏好关系是连续的和拟凹的。
- 在存在纳什均衡的严格竞争博弈中,一组行动是纳什均衡当且仅当每个行动都是最大最小化行动。
- (混合策略纳什均衡的存在性)每一个有限博弈都有一个混合战略纳什均衡。
- 一个混合策略组合是一个有限博弈的混合策略纳什均衡当且仅当每个局中人对于在其均衡策略集合上的所有行动都是无差异的。
- 若一个可转移支付联盟博弈存在非空核,当且仅当它是平衡的。
- 满足平衡贡献性质的唯一值是 Shapley 值。
- Shapley 值是唯一满足 SYM、DUM 和 ADD 公理的值。

参考文献

[1] [美]罗杰·B. 迈尔森. 博弈论——矛盾冲突分析. 于寅,译. 北京：中国经济出版社,2001.

[2] von Neumann J, Morgenstern O. Theory of games and economic behavior. Princeton, NJ: Princeton University Press, 1944.

[3] Savage LJ. The foundations of statistics. 2nd ed. New York: Dover Publications, 1972.

[4] Zermelo E. On an application of set theory to the theory of the game of chess. Translated in Schwalbe and Walker, 1913. 133-136.

[5] Kakutani S. A generalization of Brouwer's fixed point theorem. Duke Math. J., 1941, 8(3): 457-459.

[6] Kats A, Thisse JF. Unilaterally competitive games. Int'l Journal of Game Theory, 1992, 21: 291-299.

[7] Shapley LS. On balanced sets and cores. Naval Research Logistics Quarterly, 1967, 14: 453-460.

[8] Shapley LS. A value for n-person games. In: Kuhn HW, Tucker AW, eds. Contributions to the Theory of Games, volume Ⅱ, Annals of Mathematical Studies. Princeton, NJ: Princeton University Press, 1953, 28: 307-317.

第3章 演化博弈

3.1 概述

传统博弈论是在假设个体是完全理性(perfect rationality)的条件下研究博弈的均衡问题。然而,在现实世界中,个体受限于复杂的环境、自身认识、情感等因素的影响,并不能做出完全理性的选择,个体的理性是有局限的(bounded)。与传统博弈理论不同,演化博弈理论并不要求参与人是完全理性的。"有限理性(bounded rationality)"的概念是由赫伯特·西蒙(H. A. Simon)在研究决策问题时提出来的,他使用有限理性的"管理人"替代完全理性的"经济人",大大拓展了决策理论的研究领域,产生了有限理性决策理论。有限理性是指介于完全理性和非完全理性之间的、在一定限制下的理性。人的有限理性产生主要是由信息的复杂性、不确定性以及人的感知认识能力的限制引起的。演化博弈论是把博弈理论分析和动态演化过程分析结合起来的一种理论。不同于博弈论将重点放在静态均衡和比较静态均衡上,它强调的是一种动态的均衡。

演化博弈论(evolutionary game theory)起源于对生态现象的解释。它是将生物进化理论中的自然选择机制和遗传变异机制引入博弈论后的产物。当博弈论在经济学领域中广泛运用时,生物学家从中得到启示,尝试运用博弈论中策略互动思想,建构各种生物竞争演化模型,包括动物竞争、性别分配以及植物的成长和发展等。比如 1961 年 Lewontin 开始使用演化博弈论的思想研究生态问题[1]。生物学家 Smith 和 Price 使用博弈论来解释生物的进化现象。他们认为:①生物在行动中是不可能具有理性的,它们的行动是由遗传基因决定的;②在博弈中,生物的支付使用适应度来表示,即它们的繁殖能力、子孙数量;③博弈对手不像传统博弈论那样是确定的、具体的,而是随机的,参与者只知道策略分布,即种群中采用某种策略的参与者占总人口的比例;④博弈不是一次性的,而是无限重复进行的;⑤在重复进行的博弈中,有可能发生突变(mutation),即个体有可能放弃按照博弈规则得

出的最佳策略，而采取其他策略。基于以上认识，1973 年，Smith 和 Price 结合生物进化论和经典博弈论，提出了演化博弈论的基本均衡概念——演化稳定策略（evolutionarily stable strategy，简称 ESS）[2]，这标志着演化博弈论正式诞生。1978 年，Taylor 和 Jonker 在考察生态进化现象时首次提出了演化博弈论中的基本动态概念——复制动态（replicator dynamics）[3]。生物学家根据生物演化的自身规律，将传统博弈论中支付函数转化为生物适应度函数（fitness function）、引入突变机制将传统的纳什均衡精炼为演化稳定均衡以及引入选择机制构建复制者动态模型。通过对传统博弈论的一系列改造，演化博弈理论逐渐形成并发展完善。随着演化博弈论对传统博弈论的拓展，经济学家通过借鉴生物进化的思想，将演化博弈运用到经济学中，进一步推动了演化博弈的发展，包括从演化稳定均衡发展到随机稳定均衡，从确定性的复制者动态模型发展为随机的个体学习动态模型等。如今，经济学家们运用演化博弈论分析社会习惯、规范、制度或体制形成的影响因素以及解释其形成过程，也取得了令人瞩目的成绩。演化博弈论目前成为演化经济学的一个重要分析手段，并逐渐发展成为经济学中的一个新领域。

演化博弈论与传统博弈论的不同主要体现在以下 3 个关键概念上内涵的不同[4]。

（1）策略的内涵不同　在传统博弈论中，个体有特定的策略空间，个体博弈时所使用的策略来自其策略空间。而在生物系统中，个体的策略由物种的表现型（phenotype）来表达，并没有明确的策略空间。个体在继承其父辈的表现型时可能会产生突变，这些突变决定了他与其他个体在竞争中所占据的优势，或适应度（fitness）的大小。通常，个体的适应度可以使用其后代的数量进行量化。至此，经典博弈论中完全理性的观点被适者生存所取代。

（2）均衡的意义不同　演化稳定策略是演化博弈论中重要的基本概念，它与经典博弈论中的纳什均衡密切相关。演化稳定策略是指在一个种群（population）中，若所有的个体都使用了某种策略，而其他任何小的突变策略都不能入侵该种群的话，则该策略就被认为是演化稳定的。一个演化稳定策略一定是一个纳什均衡，但是只有严格的纳什均衡才是演化稳定策略。

（3）个体相互作用的内涵不同　在经典博弈中，个体之间相互作用一次，或者与相同的对手多次重复作用。而在演化博弈中，个体被随机选择进行相互作用，博弈是多次发生的且再次碰到相同对手的概率很小，避免了经典博弈论中个体的记忆效应。这种随机相互作用的限制在规则网络和复杂网络上得到放宽，网络上的演化博弈得到大量的研究。

由于与传统博弈论在以上 3 个关键概念上内涵的不同，演化博弈论将静态的决策过程赋予了动态演化的思想。

3.2　演化博弈基础

3.2.1　演化博弈的基本结构

演化博弈是关于有限理性个体在无限重复博弈中动态适应和学习的理论。演化博弈模型的结构主要由 3 部分组成：种群、博弈框架和演化机制。

（1）种群

种群是演化博弈的载体。通常种群由数量庞大的有限理性个体组成，个体使用的策略

均来自同一策略集。个体之间相互竞争，而种群也将竞争的个体之间的变化表现出来。

(2) 博弈框架

博弈框架主要指个体博弈的结构和规则。博弈结构通常可用博弈模型表示，如鹰鸽博弈模型、囚徒困境模型、公共品博弈模型等。博弈规则主要指博弈的次序和步骤等。演化博弈总是在特定的博弈结构和规则下进行的。与经典博弈不同的是，个体的策略不是通过理性选择获得的，而是通过某种传递机制获得。在重复进行的博弈过程中，每一轮博弈的个体都是从种群中随机选择的，个体再次相遇进行博弈的概率较低。

经典博弈中的支付函数也被转化为适应度函数。适应度是生物进化理论的核心概念，用来描述基因的繁殖能力。适应度函数则可被视为策略与适应度的映射关系。在生物进化领域，适应度函数的定义是比较精确和确定的。而在社会经济演化领域中，适应度函数的定义则相对模糊和不确定。策略的适应度不仅取决于它在博弈中获得的支付，还取决于个体的学习能力、个体间的交互模式等。通常支付函数不能简单地等同于适应度函数，它必须经过特定的改造才能转化为适应度函数。但是，为了简化分析，许多演化博弈模型都直接将个体的博弈支付等同于适应度。

(3) 演化机制

演化机制主要分为两种形式：选择(selection)和突变。选择是指能够获得较高支付的策略在以后将被更多的个体采用，而突变是指个体以随机的方式选择不同于群体的策略(可能是能够获得高支付的策略，也可能是获得较低支付的策略)。相比选择而言，突变是一种不断试错的过程，只有好的策略才能生存下来，因此它也是一种学习与模仿的过程，具有适应性且是不断改进的。不具备演化机制的模型不能称为演化博弈模型。典型的演化动力机制模型是复制者动态，它是一种典型的基于选择机制的确定性和非线性的演化博弈模型。

演化博弈模型有如下几个特征：①以参与者群体为研究对象，分析动态的演化过程，解释群体为何达到以及如何达到目前的这一状态；②群体的演化既有选择过程也有突变过程；③经群体选择下来的行为具有一定的惯性。

3.2.2　种群博弈

种群博弈(population games)关注的是宏观层面的变化，因此对于个体类型、个体数量及交互等方面作了许多简化的假设。一般而言，种群博弈包含3部分：二人阶段博弈、可行的策略集(一般不支持混合策略)以及个体策略的更新策略或更新规则。它隐含了以下假设。

(1) 个体是有限理性的且数量非常大，即 $N\to\infty$；

(2) 所有个体都是平等的且拥有相同的收益矩阵(对称博弈)，或者所有个体形成两种不同但本质上为同质的组(非对称博弈)；

(3) 在每一阶段(轮)的博弈中，个体以均等的概率随机地与其他个体配对(对称博弈)，或者一个组中的个体与另一个组中的个体以均等的概率随机配对(非对称博弈)，也就是说，相关的社交网络是一种最简化的形式；

(4) 与博弈频率相比，策略更新是极少会出现的情况，因此，更新规则可以基于一个策略的平均成功率；

（5）所有个体均使用相同的更新策略；

（6）个体均为弱记忆体，也就是说，其遗忘因子非常小：$\delta \to 0$。

在种群博弈中，由随机配对过程、混合策略、随机更新规则等所引起的波动，可以被平均的方法消除。从这一角度来看，种群博弈也被认为是一种基于平均场的演化博弈理论。

以上这些简化的条件和假设可以在有限的状态变量上，观察种群的宏观行为属性和博弈的动态性。

3.2.3　复制者动态

演化博弈模型必须包含演化动力机制假设，即策略在种群中的更新规则。针对实际的问题，需要适当的、不同的动力机制。演化动力机制可以分为连续模型和离散模型，也可以分为确定性模型和随机模型。在这些分类中，根据具体条件，可以形式化出大量不同的更新规则。

在宏观层面，目前为止研究最多的连续演化动力机制是复制者动态。该更新规则是 Taylor 和 Jonker 在 1978 年首次提出来的，它在生物进化模型中具有特殊地位。复制者动态被假设为给定策略 i 在种群中的人均增长率 $\dot{\rho}_i/\rho_i$ 与适应度之差成正比：

$$\frac{\dot{\rho}_i}{\rho_i} = f_i - \bar{f}$$

其中，ρ_i 是采用策略 i 的个体在种群中所占的比例，f_i 是策略 i 的适应度，$\bar{f}$ 是种群的平均适应度，$\rho=(\rho_1,\rho_2,\cdots,\rho_n)$，$\sum_i \rho_i = \sum_i N_i/N=1$。适应度表示个体的演化成功程度，可表示为博弈的支付。在种群博弈中，博弈的收益矩阵为 A，策略 i 的期望收益为 $(A\rho)_i$，也是策略 i 的适应度，而平均收益表示为 $\rho \cdot A\rho$，也是平均适应度，则上式表示为

$$\dot{\rho}_i = \rho_i((A\rho)_i - \rho \cdot A\rho)$$

该式通常称为复制者动态公式的泰勒形式（Taylor form）。

在不同的假设条件下，复制者动态公式也可以采用如下形式：

$$\dot{\rho}_i = \rho_i \frac{(A\rho)_i - \rho \cdot A\rho}{\rho \cdot A\rho}$$

该式则称为梅纳德史密斯形式（Maynard Smith form）。

需要注意的是，如果在初始时不包含给定策略 i，即对于给定策略 i，有 $\rho_i(t=0)=0$，那么对于所有时刻 t，$\rho_i(t)=0$。也就是说，复制者动态不能产生新的策略。此外，复制者动态给最初作为静态概念引入的演化稳定赋予了明确的动态意义。

3.2.4　演化稳定策略

演化博弈论的一个核心问题是一个种群中策略配置的稳定性和鲁棒性。事实上，根据“间断均衡”理论，生物学上的演化不是一个连续性的进程，而是一个突发性的物种演化和灭绝的状态转移事件。在突发性状态转移事件之间，则是长期相对平静和稳定的时间。一方面，间断均衡理论能够用中间物种与分支物种的理论解释丢失的化石记录；另一方面，其暗含着这样一种理论：能够观察到的大多数生物结构和行为模式都拥有很高的稳定性。从这一点上说，一个通过博弈理论模型给出的解只是一个演化状态中的可能的稳定解。

演化稳定策略的概念最初是由 Smith 和 Price 于 1973 年在对称种群博弈的研究中提

出来的。对于一个给定的策略，如果占种群绝大多数的个体选择该给定策略时，能够抵御使用不同策略的小的突变群体的入侵，那么这个给定策略就是演化稳定策略。种群中个体更多考虑的是演化稳定策略的效用，而不是种群中的突变者，因此，长期而言，种群会驱逐变异者而保持稳定。演化稳定策略在整个演化过程中均为优势策略，因此，在真实世界中观察到的策略通常都是演化稳定策略。通常来说，演化稳定性隐含着入侵屏障（invasion barrier）的概念，演化稳定策略仅能在突变未达到临界频率，即未达到入侵屏障的情况下才可能抵御突变者入侵。需要说明的是，演化稳定策略这一概念与实际的博弈动态过程没有任何关联，因此，它是一个“静态”的概念。其唯一必须成立的假设在于一个表现更好的策略必须要有一个更高的复制（生长）率。

定义 3.1（演化稳定策略） 设一个演化博弈的博弈模型是二人对称博弈，纯策略 s 表示突变策略，如果纯策略 s^* 是演化稳定策略，则对所有 $s \neq s^*$ 有

（1）$E(s^*, s^*) > E(s, s^*)$，或

（2）$E(s^*, s^*) = E(s, s^*)$，且 $E(s^*, s) > E(s, s)$，

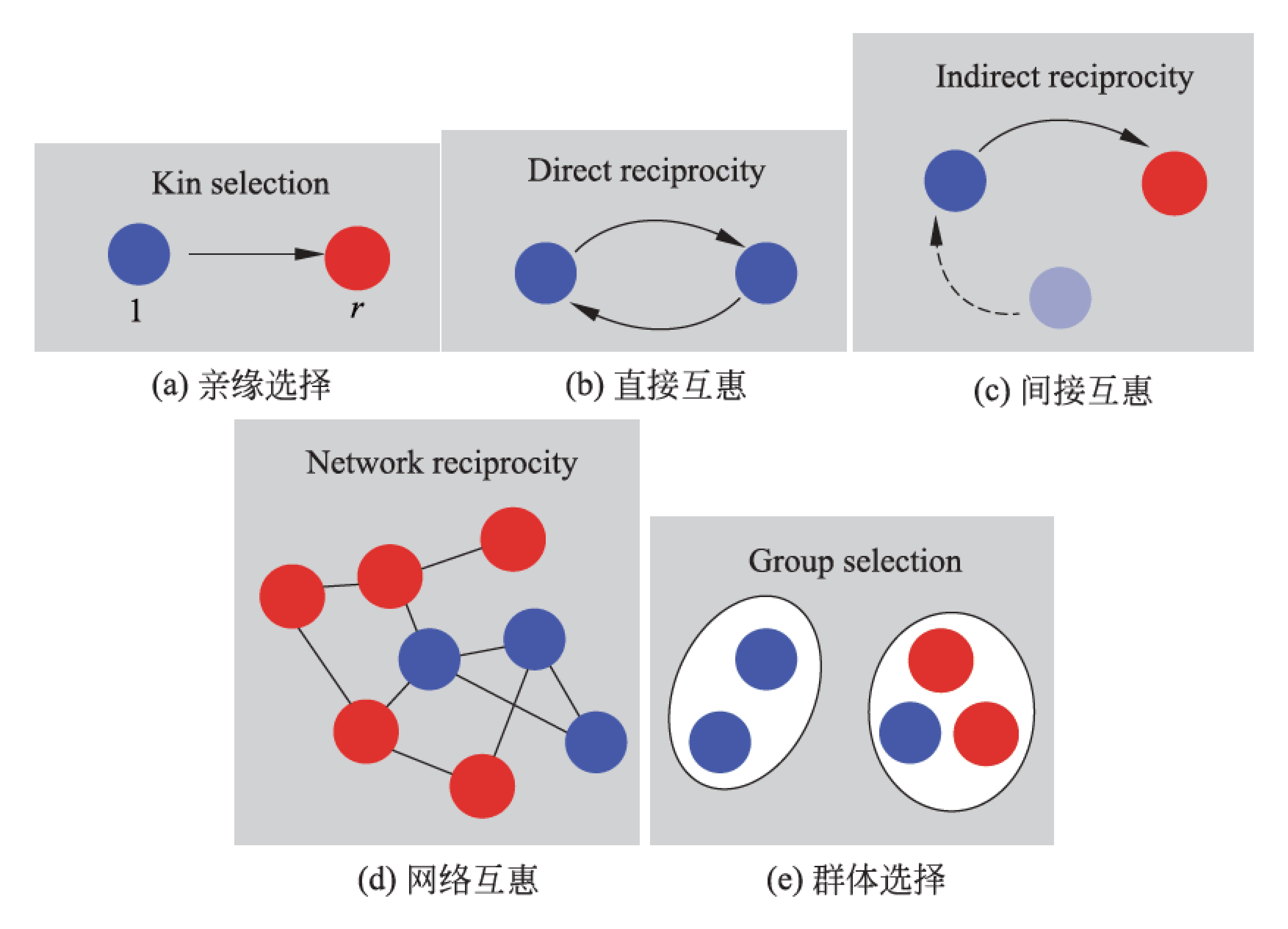
其中，$E(x, y)$ 表示策略 x 遇到策略 y 时的收益。

从上述定义中可以看出，条件（1）比条件（2）更严格，条件（1）也称为严格纳什均衡。任何满足条件（1）的演化稳定策略称为强演化稳定策略（strong ESS）。如果一个演化稳定策略是纳什均衡策略但不是严格纳什均衡策略，那么它就必须满足条件（2）。条件（2）说明，对于不满足条件（1）的演化稳定策略，如果突变策略遇到该策略的收益与该策略遇到自身的收益一样好，那么该策略遇到突变策略的收益必须严格优于突变策略遇到自身的收益。任何不满足条件（1）但满足条件（2）的演化稳定策略也称为弱演化稳定策略（weak ESS）。

演化稳定策略具有以下一些特征：首先，演化稳定策略不依赖于演化过程。它与纳什均衡一样都是静态的概念，并不探讨均衡是如何获得的。在某些情况下可以从博弈的支付矩阵中直接判断出演化稳定策略；其次，演化稳定策略一定是纳什均衡，反之，则不然。因此，演化稳定策略是纳什均衡的一种精炼。

3.3　网络演化博弈

通常认为合作行为是有着付出成本（cost）的利他行为。合作者付出一定的代价使对手或群体获得更高的效用，而不合作者不需要付出代价而获得好处，其效用比合作者更高。按照优胜劣汰、适者生存的原则，群体中低收益的合作行为会被高收益的背叛行为完全取代而消失。所有个体都选择背叛行为，最终会使得整个群体的效用最小，出现社会困境问题。但是无论在自然界中还是人类社会中都广泛存在着合作现象。如果没有特定的机制来促使合作的涌现和稳定维持，这种现象是难以解释的。合作的涌现与稳定维持的机制是演化博弈研究的重要课题和方向。2006 年，哈佛大学 Nowak 教授在 *Science* 上发表的论文中总结了促进合作产生的 5 种机制：亲缘选择（kin selection）、直接互惠（direct reciprocity）、间接互惠（indirect reciprocity）、网络互惠（network reciprocity）和群体选择（group selection），如图 3-1 所示。



(a) 亲缘选择　(b) 直接互惠　(c) 间接互惠

(d) 网络互惠　(e) 群体选择

图 3-1　5 种促进合作产生的机制[5]

网络演化博弈起始自 1992 年 Nowark 和 May 将空间结构引入演化博弈中。通常，一个完整的网络演化博弈主要包含以下 3 个要素：博弈模型、作用网络和更新规则。博弈模型通过策略和效用描述了个体间相互作用的方式；作用网络描述了个体之间的关系，决定了个体博弈的范围；更新规则则描述了个体学习和模仿的方式，反映了模型的动态演化。

目前的网络演化博弈研究主要集中在以下 3 个方面：①单一的静态网络上不同的网络结构和演化规则对合作行为的影响；②个体行为策略与个体关系结构在一定的更新规则下都进行更新的共演化；③相互依赖网络中合作的演化。基于个体的模拟仿真是网络演化博弈的主要研究手段。合作频率(f_c，frequency of cooperators)，亦称合作者比例，是用来评价系统中合作水平高低的指标，它是指系统中使用合作策略的个体的比例。

网络演化博弈的一般流程是：

(1) 初始时按相同的概率将网络中的个体分配为合作者($s_i=C$)或背叛者($s_j=D$)；

(2) 个体按照博弈规则与其邻居个体进行博弈并计算他们的效用u_i(个体与其所有邻居个体博弈的收益之和)；

(3) 个体按照一定概率 $w(s_i \leftarrow s_j)$学习或模仿另一个个体的策略；

(4) 不断重复步骤(2)和步骤(3)，直至系统达到稳态。

个体 i 和个体 j 是随机选择的，并且更新操作通常是随机顺序(random sequential)进行的。策略的更新所使用的概率 $w(s_i \leftarrow s_j)$有多种不同的定义方式，最常见的定义方式是费米函数(Fermi function)：

$$w(s_i \leftarrow s_j) = \frac{1}{1+\exp[\beta(u_i - u_j)]}$$

这里，β 表示非理性选择的噪声，也称为选择强度(intensity of selection)。当 $\beta \to 0$ 时，i 以 0.5 的概率保持当前策略，以 0.5 的概率模仿 j 的策略；当 $\beta \to \infty$时，个体趋于完全理性，即只有当$u_j > u_i$时 i 才会进行模仿 j 的策略；当 $\beta > 0$ 时，即使$u_j < u_i$，i 仍会以一定的概率模仿 j 的策略。

另一个常用的概率定义方式如下式所示，它适用于异质网络。

$$w(s_i \leftarrow s_j) = \frac{u_j - u_i}{\Delta k_>}$$

这里，$k_>$是k_i与k_j中大的那个节点度数，对于囚徒困境，$\Delta = T - S$，对于雪堆博弈，$\Delta = T - P$，而对于猎鹿博弈，则有 $\Delta = R - S$。

不同于以上两种更新方式，还有一种常用的确定性更新方式——best-take-over。在这种策略更新方式下，个体总是学习效用最高的策略。

3.3.1 静态网络上的演化博弈

在 Nowak 和 May 引入空间结构之后，最初研究的网络结构集中在规则网络上[6]。这些研究发现网络结构的引入通常情况下能够促进合作的演化，但也发现，在某些情况下，网络结构也未必有利于合作的演化[7]。随着复杂网络研究的兴起以及无标度网络特性在现实世界广泛存在这一现象的发现，网络演化博弈中的网络结构也由规则网络扩展到了复杂网络上。

静态网络互惠是网络演化博弈研究的主要发现[6]。在静态网络互惠的支持下，合作者通过集聚成团相互扶持来抵御背叛者的入侵，同时处于合作者集聚团边缘的个体能够以优势地位吸引背叛者的加入，从而实现团聚的扩展，其原理图如图 3-2 所示。图 3-2 显示的是社交网络上基于囚徒困境的演化博弈中合作演化的时间进程图。从图中可以看到，合作演化的时间进程可以按顺序分为两个部分：忍耐期（enduring，简称 END）和扩张期（expanding，简称 EXP）。在合作演化的忍耐期，开始时，合作者迅速被背叛者侵占，导致只有小部分的合作者通过集聚成紧凑的团聚来相互扶持，抵御背叛者的入侵；而在扩展期，处于团聚边缘的合作者吸引其邻居个体中的背叛者加入团聚，实现了团聚的扩张。这种合作者比例“先下后上（first-down-then-rise）”的现象是静态网络互惠的著名标签。

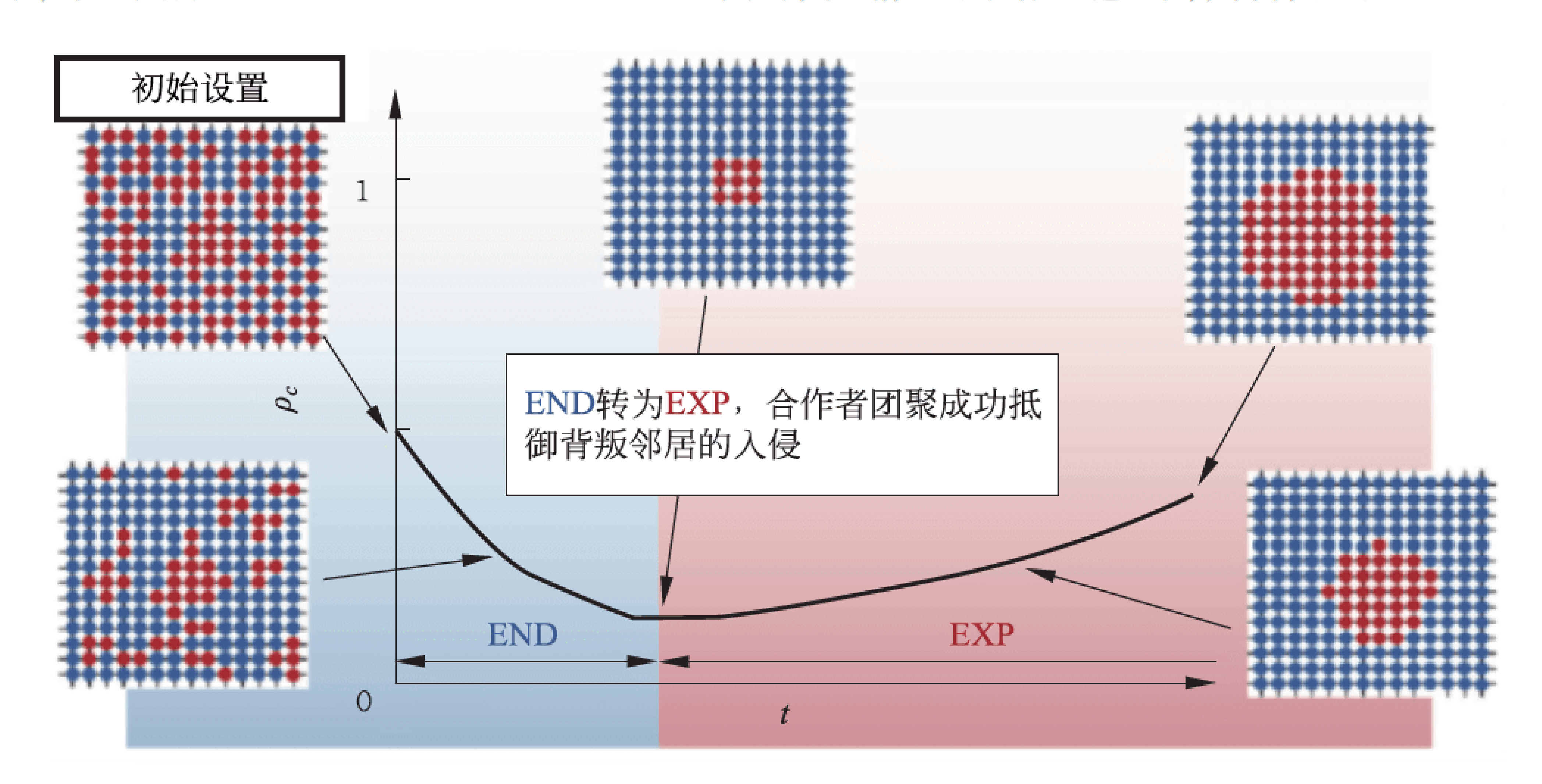


图 3-2　静态网络互惠随时间演化原理图

红色表示合作者，蓝色表示背叛者[11]

由于无标度网络中存在的高度异质性，Santos 和 Pacheco 将这种异质性引入到演化规则中，提出了新的策略更新规则。为了进一步弄清这种异质性的影响，Santos 等人对网络的异质性的影响进行了对比研究，如图 3-3 所示。他们在囚徒困境、雪堆博弈和猎鹿博弈 3

种博弈下对比了 4 种不同网络模型下的异质度的影响：完全网络(complete)、度分布呈高斯分布的单标度网络(single-scale)、基于随机连边的无标度网络和基于 BA 模型的无标度网络。

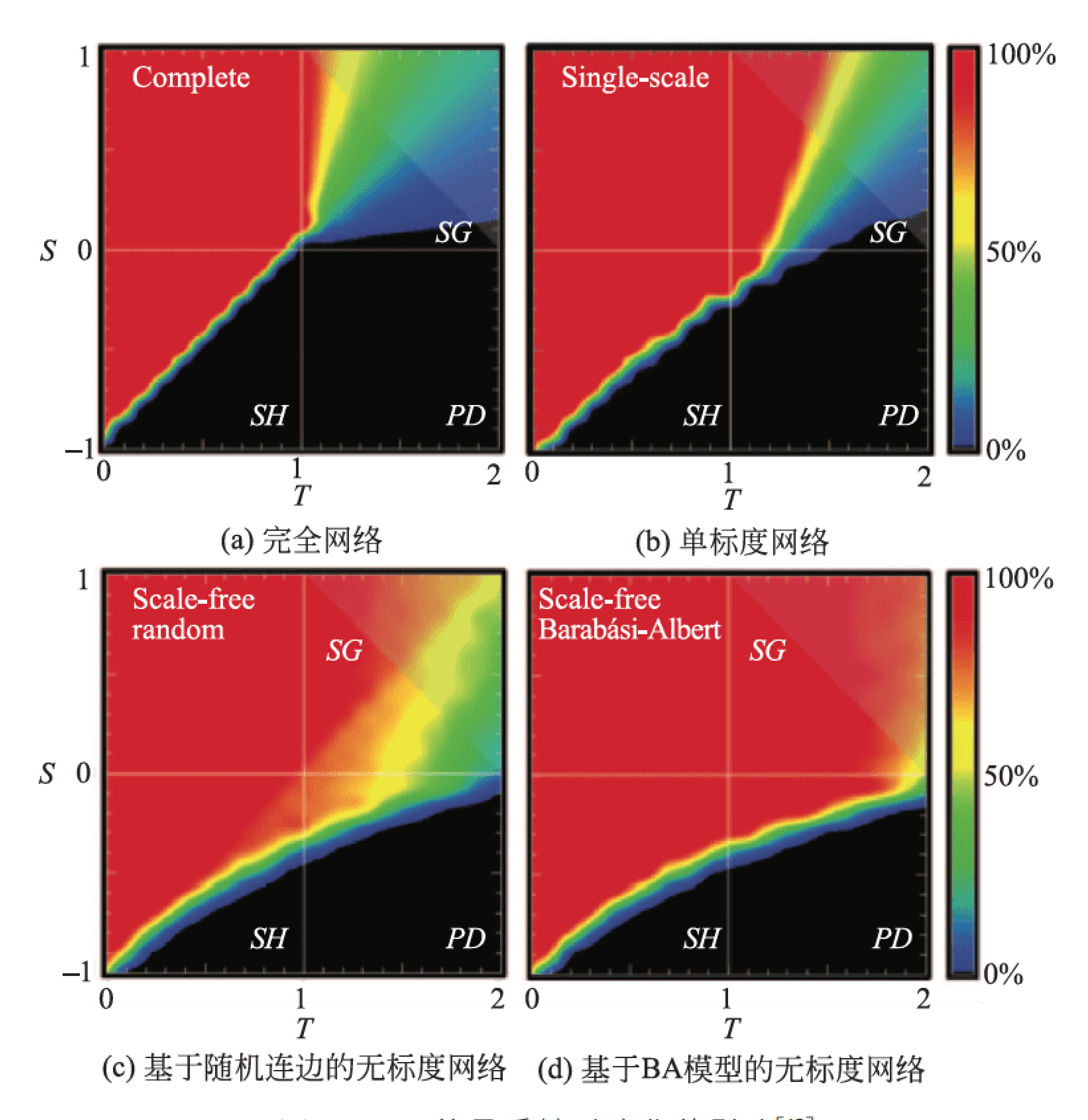


图 3-3　网络异质性对演化的影响[12]

研究结果显示，异质网络比同质网络更容易维持合作，这种现象与采用哪种博弈模型无关。从图 3-3 可以看到，网络的异质性提供了促使合作涌现的有力机制，即使这种异质性并不十分显著，但它仍能对合作的演化带来可观的影响。

在单标度网络和无标度网络中，在 3 种博弈模型下，网络中的合作水平都得到了整体性的提高。同时，大度数节点的直接相连能够使网络个体承受更大的付出以维持合作，显示出合作的维持与个体之间复杂的连接关系有关。

社区结构作为社交网络中普遍存在的、重要的中观结构，其对合作演化的影响也得到研究。通常认为，社区在社交网络中属于内部连接紧密，外部连接稀疏。Chen 等人研究了带社区结构的无标度网络对合作演化的影响[13]，该网络特性采用偏好连接的方法生成。研究结果显示，由于社区结构的存在，无标度网络的高度异质性并不总是会促进合作。这种对合作的抑制现象在 ER 随机网络和社区网络的对比研究中也有发现[14]。产生这种现象的原因是，处于不同社区的合作者可能并不是直接相连的，而他们也无法通过社区间的连接相互支持。这导致了合作者的团聚很难成功形成，整个网络中合作者无法取得优势。但是，当社区内部的连接密度增加时，合作者团聚成功形成的可能性也有所增加，合作得到促进。Wang 等人则研究了局部社区和全局社区并存的多级社区结构——层次社区结构的影响[15]。一个社区内部的交互使用公共品博弈模型作为隐喻(metaphor)，这样，层次社区结

构也就构成了一个多级公共品博弈（multilevel public goods game，简称 MPGG）。当社区规模和社区数目足够大时，社区内部的弱模仿或者社区之间的强模仿都能够促进合作。这种现象出现的原因在于社区内部的弱模仿可能会导致合作者的分级，而社区间的强模仿会使得合作者团聚一旦出现便马上流行开来。由以上研究可以明显看到，宏观特性相似的社交网络，由于它们的中观结构特性的不同会带来非常不同的结果[16]。在具有社区结构的社交网络中，不仅是中心个体(hub)，中心个体之间的直接连边也在演化中扮演着重要的角色。

在社交网络中，由于属性的不同，个体的行为表现也会不同。这种属性的不同也被称为多样性（diversity）。Santo 等人引入了个体社会多样性，这种多样性体现在个体参与的公共品博弈的规模和数量及其在每个博弈中的贡献[17]。Shi 等人则提出了群体多样性，这种多样性体现在公共品博弈的不同放大因子所带来的群体异质性[18]。而 Du 等人则研究了个体多样性带来的非对称付出成本的影响，即在博弈双方共同合作中存在的不平等现象[19]。对于共同合作，合作者 i 与 j 付出的成本分别是 $r\Lambda_i$ 和 $r\Lambda_j$，其中，r 是付出的成本，则其非对称系数分别定义为

$$\begin{cases} \Lambda_i = \dfrac{k_j^\alpha}{k_i^\alpha + k_j^\alpha} \\ \Lambda_j = \dfrac{k_i^\alpha}{k_i^\alpha + k_j^\alpha} \end{cases}$$

这里，α 是条件参数，它可以调整非对称成本的分布。如果 $\alpha > 0$，则高度数节点剥削低度数节点，而如果 $\alpha < 0$，则情况正好相反。非对称的付出成本增强了异质财富分布在决策过程中的领导作用，不仅带来了较高的合作密度，也稳定了合作行为。

影响力（influence）作为社交网络中个体的重要的社会属性，通常被认为是一种策略迁移或教导的能力。这种能力能够使个体更容易把自己的策略传递给其他个体，即有影响力的个体的策略更容易被其他个体模仿。Szolnoki 和 Perc 引入了影响力因子来表示影响强度[20]。当个体 i 成功地将其策略传递给个体 j 时，其影响力因子就会增加一个正数 $\Delta w \ll 1$，使得 $w_i \to w_i + \Delta w$。而在文献[21]中，个体被分为两种类型：A 型（影响力个体）和 B 型（无影响力个体）。Du 等人则定义了一种与度相关的影响力[22]，高度数的个体有着高影响力。当一个个体决定更新其策略时，他会从其邻居中按照影响力的高低去选择个体进行模仿。这样，高影响力个体的策略有更多的机会被其他个体模仿。在文献[23]中，当一个个体进行策略更新时，他根据概率从其邻居中选择一个个体进行模仿，这个概率由被选个体的影响力在这些邻居个体中所占的比例决定。如果策略模仿成功进行，那么被选个体的影响力就会增加。以上这些研究中，不论哪种影响力的定义，个体在影响力的多样性上都促进了社交网络上合作的产生。

其他形式的个体的多样性，比如声誉及其容忍度、占有资源多样性、年龄等在静态网络结构中的影响也得以研究。

3.3.2　动态网络上的演化博弈

无论是在自然界还是人类社会中，个体间的关系并不是一成不变的，会随着个体的交互行为、兴趣等因素而发生变化。个体通过调整交互作用的对象来获取更高的收益或改善自身的社会属性。

个体策略和关系的共演化对网络的合作水平有重要的影响。根据连边变化对个体效用影响的方式,共演化规则主要可分为 3 类:①依据连边的策略类型,使用"活跃连接(active linking)"[25]确定的概率来创建或切断连边,如图 3-4 所示;②策略敌对的连边的切断概率依赖于相关个体的效用,而获胜个体切断与失败个体的连边之后,与失败个体的一个邻居建立新的连边,如图 3-4(b)所示;③每当个体策略发生更新时,除了保留与被模仿个体之间的连边以外,其他连边均被切断,如图 3-4(c)所示。

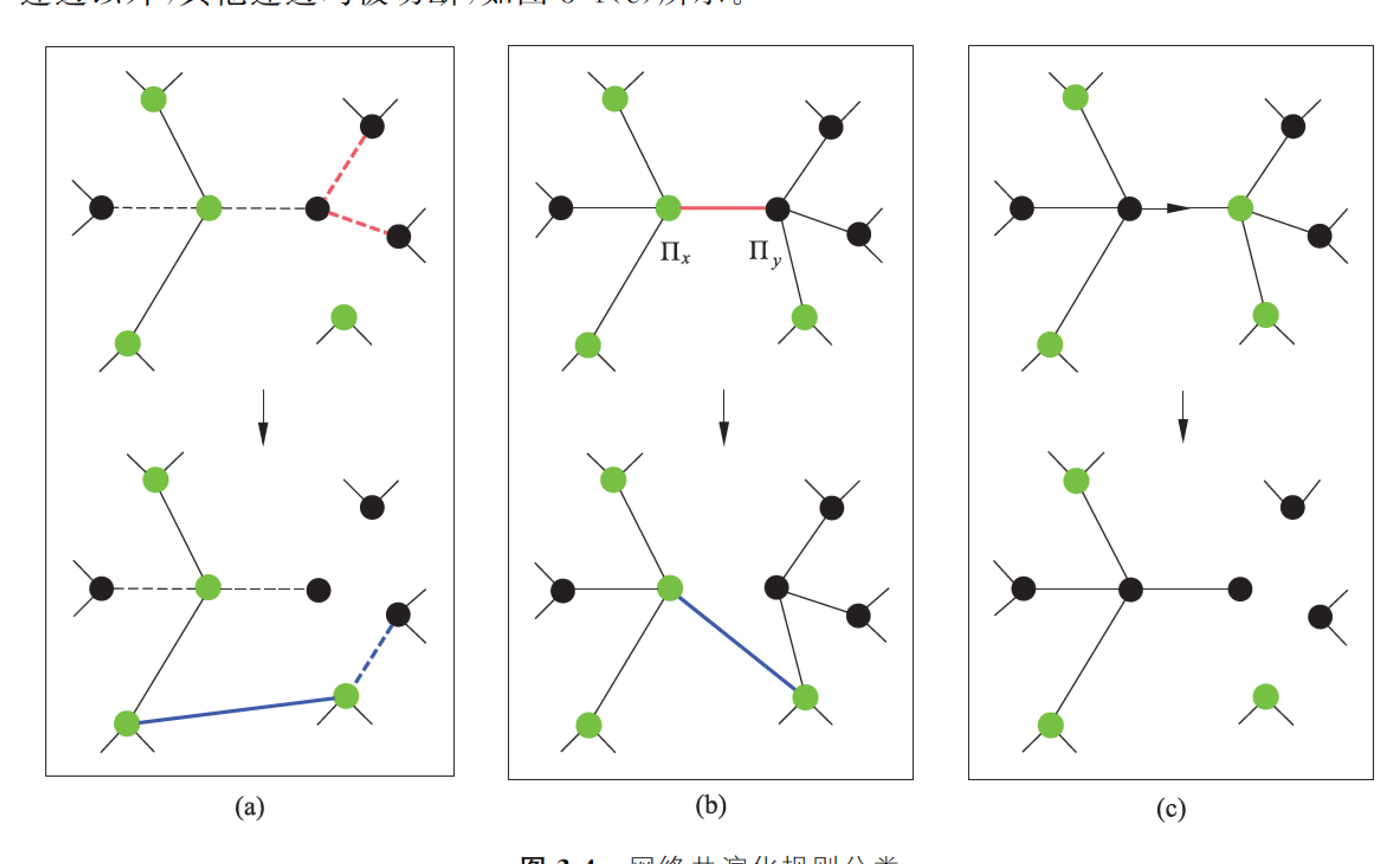


图 3-4　网络共演化规则分类

绿色表示合作者,黑色表示背叛者;红色边表示被删除的边,蓝色边表示被创建的边[26]

从个体角度来看,个体在调整其关系时总会根据一些因素来切断和创建新的连边,这些影响因素包括效用、社会属性等。个体为了获得更高效用而调整其博弈对手是个体调整关系的一个比较直观的目标。Zimmerman 等人研究了个体切断提供最小效用连边的情况[27]。在实际中可以观察到,D-D 连边所产生的效用总是最小,这导致了 D-D 连边的两个个体会选择切断原有的不利连接并寻找其他个体建立新的连接。而在 Santos 等人提出的连边调整模型中,对于一条连边e_{ij},如果 j 是合作者,那么 i 对于该连边是满意的,否则,i 对该连边不满意。同样地,j 也会按照相同的方式来评估对e_{ij}的满意度。如果两者都对该连边满意,那么e_{ij}会维持原状,否则,两者会竞争重新连边的主动权。而重新连边的主动权则与两者的效用密切相关。同时,在他们的模型中对策略更新和连边更新引入了不同的时间尺度,即策略更新和连边更新的频率不同。τ_s表示策略更新的时间尺度,τ_e表示连边更新尺度,$W=\frac{\tau_s}{\tau_e}$表示两个时间尺度的比值。$W\to 0$ 时,连边更新的频率趋于 0,网络也变成静态结构。随着 W 的增大,连边更新的频率也增大。当 W 大于特定的阈值后,合作者能够将背叛者消灭,网络结构的异质性也达到最大。

值得注意的是,演化生成的网络的异质性所呈现的模式与社交网络中观察到的模式是

吻合的[8]。Pacheco 等人提出了“活跃连接(active linking)”机制[25]。在该机制中,个体间的关系以一定比率建立并且具有生命周期(life-time)。当活跃连接机制的频率远大于策略更新的频率时会引起收益矩阵发生改变,如囚徒困境博弈转化为协调博弈,雪堆博弈转化为完美博弈。更进一步地,van Segbroeck 等人[28]在活跃连接机制中引入了个体切断连边率 γ。$\gamma \to 1$ 的个体会以非常高的频率切断不利的连边,而 $\gamma \to 0$ 的个体则倾向于极少切断连边。实验结果显示,这种个体处理连边方式的多样性有助于提高网络的合作水平。Fu 等人研究共同合作连接需要双方共同建立,而单方合作则可以单方面切断的情况,即只有合作者被允许切断与背叛者之间的连边并随机选择新的个体建立连接[29]。结果发现,存在一个策略更新频率和连边更新频率的最优比例,使得网络的合作水平最好。此外,在群体交互下,个体对于效用的偏好的影响也得到了研究。

除了连边对手的策略、效用等直接因素,个体的社会属性也是个体在连边调整时需要考虑的因素。在社交网络中,声誉(reputation)是个体的一种常见的和重要的社会属性。个体的声誉反映了个体在其行为历史中合作的程度。Nowak 等人在 1998 年就在网络演化博弈中使用了“印象分(image score)”的概念来反映个体的声誉[30]。Fu 等人在连边调整机制中引入了声誉机制,这里所说的声誉的定义与印象分的定义类似[31]。当个体进行连边调整时,他切断了与邻居中最低声誉个体的连接,并依概率 p_r 与其邻居中声誉最高的节点建立新连接,或者以概率 $1-p_r$ 从网络中随机选择个体建立新连接。演化生成的网络的累积度分布呈幂律分布,表现出高度的异质性。高度数的节点更容易成为有着高声誉度的合作者。声誉与度的正相关性提供了促进合作涌现的正反馈机制。此后,基于声誉的连边机制在社交网络演化的不同方面得到研究,比如社交连接与地理连接共存、有向社交网络中的非排他公共品博弈、昂贵的声誉建立过程,以及声誉的记忆效应等。

Wang 等人研究了公共品博弈下基于期望(expectation)的群体形成机制[32]。这里,期望是指个体对其所处的群体环境的满意程度。如果一个群体无法满足个体的预期,就会与主持该公共品博弈的个体断开连接从而退出该群体,同时与其他个体建立连接。因此,期望不仅能够控制连边调整的趋势程度,也决定了群体的生存能力。实验结果发现,期望水平存在着最优和最差的区间,它们分别对应了最好和最差的合作水平。这个最优的期望区间是期望处于中等水平,此时演化生成的网络具有相当高的异质性。

Zhong 等人则在连边更新机制中引入了信任(trust)[33]这一概念。Szolnoki 等人研究了策略更新依赖的共演化机制[34],该机制与图 3-4(c)所示的规则一致。而在他们的另一工作[34]中则研究成功个体比不成功个体应当拥有更多社交关系的情况,即每当个体成功地将其策略传递给其他个体时,他的度就会通过与新个体建立社交关系的方式增加。此外,Wu 等人研究了以随机概率断开一个连边而其中的一个个体又获得建立新连边的情况[35]。动态连边机制能够促进合作的涌现这一现象得到了实证研究的支持。动态连边频率的影响和有代价的连边调整也得到了实证研究。

3.3.3 相互依赖网络上的演化博弈

在现实世界中,由于生活、工作、兴趣等多方面因素,个体会同时身处多个群体,甚至是多个社交网络。比如,一个人在使用社交网络服务时,由于其信息行为的内容、兴趣爱好等因素而属于不同的群体。他可能会加入多个微信群或 QQ 群,或者处于微信的朋友圈,也会

处于与微信联系紧密的但不同的 QQ 好友圈。此外，他还可能同时使用了新浪微博以及人人网、LinkedIn 等社交网络服务。因此，现实社会中不同的群体或社交网络之间是会相互关联、相互连接的。这些彼此相互联系的群体、社交网络等构成了“网络的网络（networks of networks）”，如图 3-5 所示[36]。

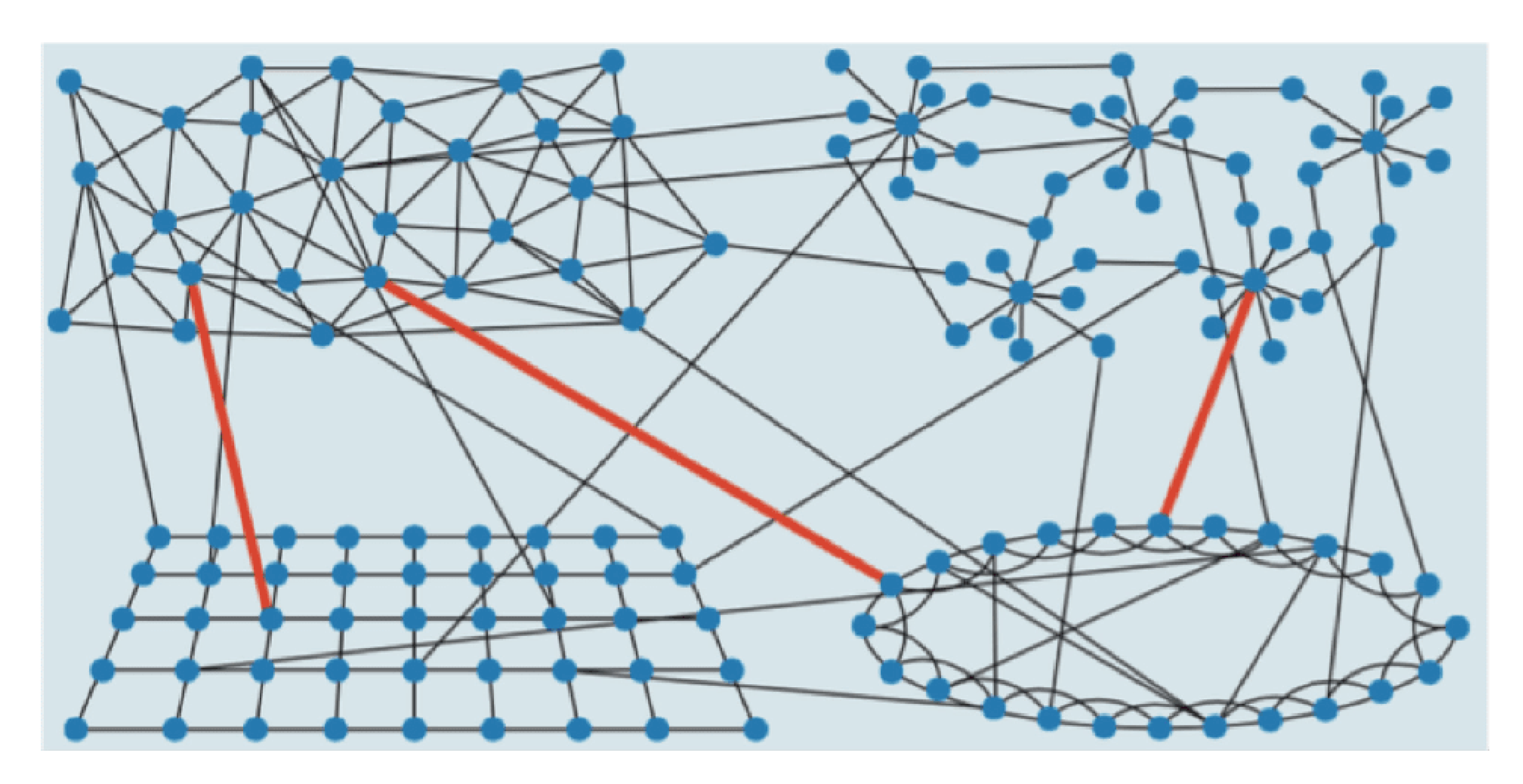


图 3-5　网络的网络

在现实中的网络通常都是相互连接的。虽然目前还不知道这种网络的一般形态，但是一些量化形式仍然通用。网络间的连接至少有两种类型：相连（connetivity，与单一网络中的连接相似，图 3-5 中用黑色连边表示）和依赖（dependency，对节点来说生死攸关的连接，图 3-5 中用橙色表示）[36]。

在这两种类型的网络间连接中，依赖连接的作用与影响得到了研究者们的更多关注，这种网络也被称为相互依赖网络（interdependent network）。相互依赖网络的经典案例是 2003 年 9 月 28 日发生在意大利的大停电事件。在该事件中，由于一个电站的失效导致依赖于该电站电力供应的一些互联网节点的失效，而这些互联网节点的失效又导致调度系统无法对电网进行有效的调控，从而诱发了更大规模的电力节点的失效，如此诱发的级联崩溃效应最终导致了大停电事件，如图 3-6 所示。2010 年，Buldyrev 等人在 *Nature* 上首次发表

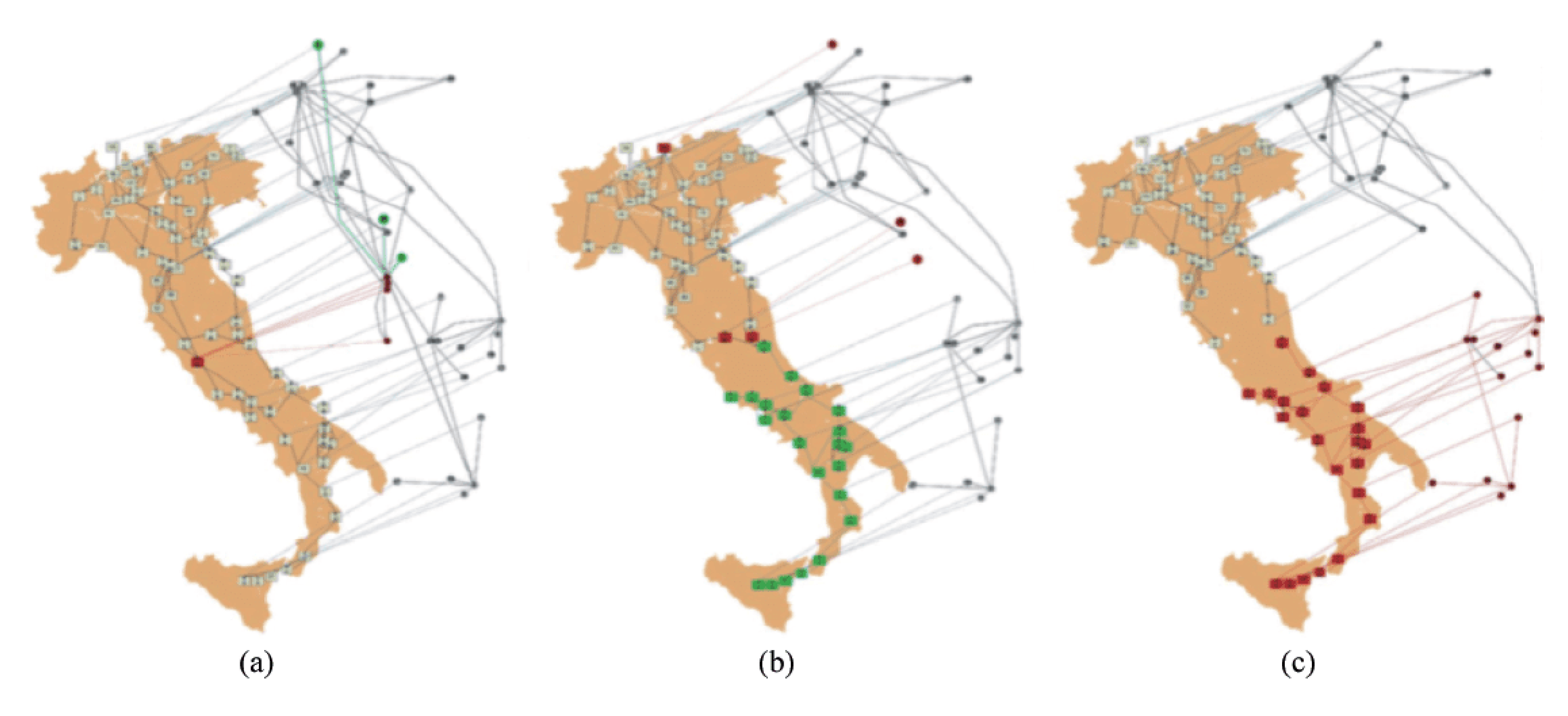


图 3-6　意大利大停电事件中的级联失效过程

了研究相互依赖网络鲁棒性的文章[37]，提出了一个简单的相互依赖网络级联失效模型，刻画了相互依赖网络级联失效的过程，发现了相互依赖网络从连通态(有序)到破碎态(无序)是一个非连续的相变过程。该文发表后获得了广泛的关注，相互依赖网因此得到了大量的研究。

图 3-6 中的数据源自真实的数据，意大利地图上显示的是电力节点与电力网络，地图的右上方显示的是互联网及其节点。其中，图 3-6(a)所示为电力网络的一个电站失效(地图上的红色节点)导致依赖于该电站的互联网服务器节点(地图右上方网络上的红色节点)失效。这些失效的互联网节点将导致与其相连的一些互联网节点失效(地图右上方网络上的绿色节点)。图 3-6(b)所示的这些新失效的互联网节点(地图右上方网络上的红色节点)导致依赖于这些互联网节点的电站也失效(地图上的红色节点)。地图上的绿色节点表示在下一步中将失效的电力节点。如图 3-6(c)所示，新失效的电力节点(地图上的红色节点)导致依赖于它们的互联网节点失效(地图右上方网络上的红色节点)[37]。

相互依赖网络研究的兴起，大大拓展了网络演化博弈的研究范围。相互依赖网络中网络间的依赖关系，从形式上可以分为有连边和无连边，而从影响的作用方式上则可分为：直接影响个体的效用、影响策略更新以及同时影响效用和策略更新。目前相互依赖网络上的演化博弈的研究中网络结构以方格网络为主。Wang 等人研究了相互依赖网络的依赖对个体效用函数的影响是非对称的情况，即两个网络对彼此的影响是不一样的[38]。假设个体 i 在网络 A 中，网络 B 中对应的个体是 i'，则依赖关系通过如下效用函数表达：

$$\begin{cases} u_i = \alpha\pi_i + (1-\alpha)\pi_{i'} \\ u_{i'} = (1-\alpha)\pi_{i'} + \alpha\pi_i \end{cases}$$

这里，π_i和$\pi_{i'}$分别表示 i 和 i'在各自网络中的收益，α 则表示两个网络之间收益影响的偏差。实验结果发现，这种相互依赖网络的效用的非对称影响能够显著促进系统的合作水平，即两个网络中合作的平均水平比单一网络的合作水平要高。与非对称效用函数不同，Wang 等人定义了网络对称的效用函数[40]

$$u_i = (1-\alpha-\beta)\pi_i + \alpha\langle\pi_i\rangle + \beta\langle\pi_{i'}\rangle$$

这里，π_i是 i 在一个网络中的收益，$\pi_{i'}$是 i 在另一网络中的对应个体 i'的收益，$\langle\pi_i\rangle$和$\langle\pi_{i'}\rangle$则分别是 i 和 i'各自所有邻居的平均收益。基于这种网络对称定义的效用函数，相互依赖网络互惠(interdependent network reciprocity)自发地涌现。如果相关联的合作者团聚能够在两个网络中同时形成，那么合作涌现的促进程度将比传统单一网络互惠大得多。另一方面，如果这种合作者团聚协调形成被干扰，如在一个网络中合作者团聚形成而另一个网络中的合作者团聚没有形成，那么这种网络互惠的机制在两个网络中都会完全失效。如图 3-7 所示。

图 3-7(a)～图 3-7(d)显示的是一个网络的合作演化，图 3-7(e)～图 3-7(h)显示的是另一个网络的合作演化。蓝色表示合作者，红色表示背叛者。初始时，各有两个合作者的小团聚，如图 3-7(a)和图 3-7(e)所示。图 3-7(a)和图 3-7(e)各自左下角的合作者团聚在演化一开始就可以利用相互依赖网络互惠机制，都成功生存并扩散至各自的网络[40]。

Wang 等人研究了网络间的最优相互依赖[39]。在他们的模型中，只有部分个体能够与其另一网络中对应的个体建立连接，这部分个体称为杰出个体(distinguished)。杰出个体的效用受到另一网络中对应个体的影响，$u_i = \pi_i + \alpha\pi_{i'}$，$0 \leqslant \alpha \leqslant 1$。没有网间依赖连接个体称

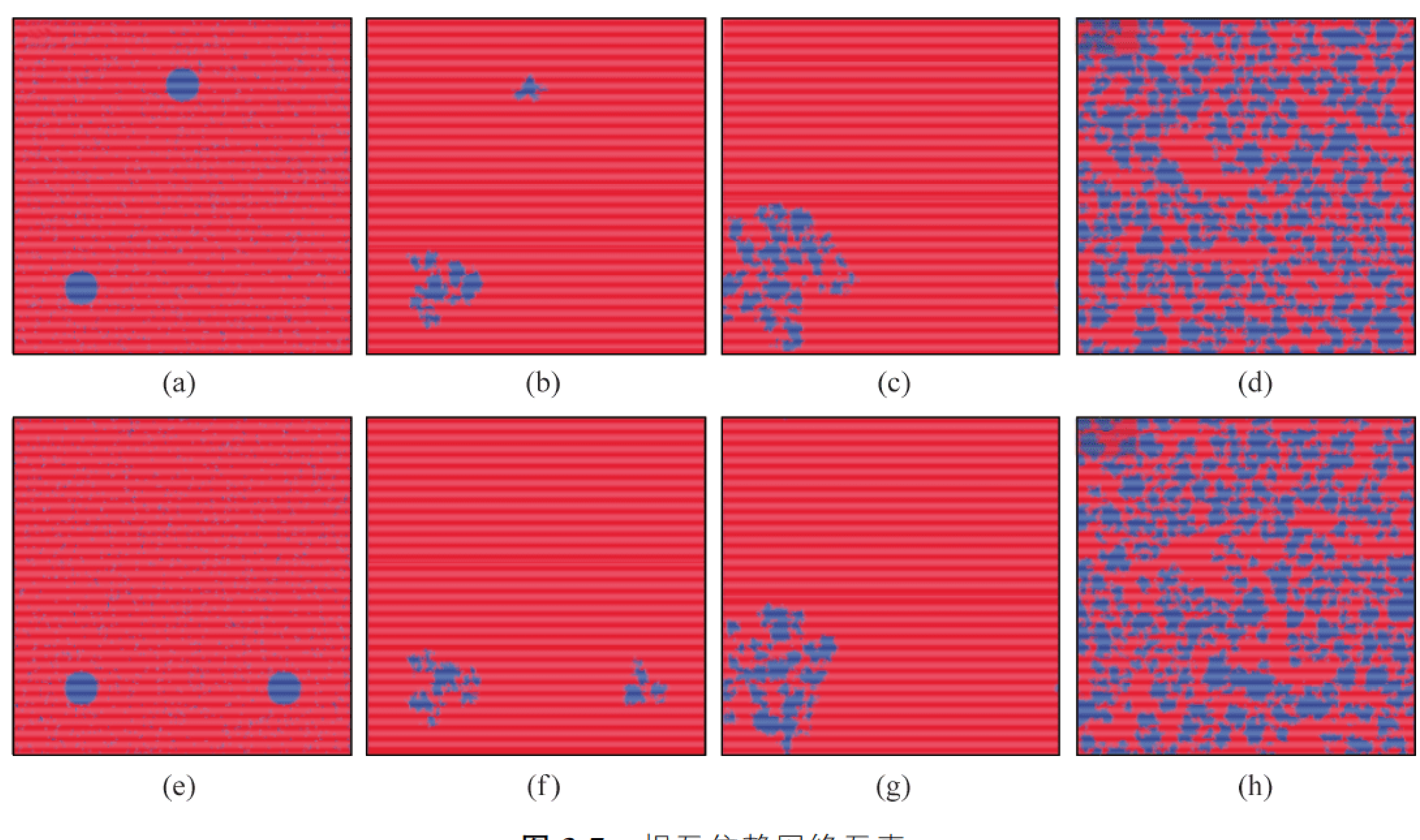


图 3-7　相互依赖网络互惠

为平凡个体(ordinary),他们的效用则不受另一网络的影响。这种个体获得效用的异质性导致了不对称的策略流动,这促使了影响力节点的涌现。这些影响力节点在各自的网络上成为强大的合作中心节点。如果网络的依赖是双向的,这种机制的表现最好。合作者团聚在两个网络上涌现并通过依赖关系相互支持。当两个网络的依赖处于中等时,相互依赖网络呈现出最佳的合作水平,如图 3-8 所示。这一现象在文献[41]的工作中亦有显现。上述结果对不同的策略更新机制(如费米规则、best-take-over 规则和比例规则)、网络结构(如聚类系数为 0 的方格网络和具有高聚类系数的三角格子网络),以及博弈模型(如囚徒困境博弈和雪堆博弈)等都具有鲁棒性。

Szolnoki 和 Perc 把相互依赖关系引入到个体策略更新机制中,即一个网络中的个体在更新其策略时会受到其在另一个网络中的对应个体的策略信息的影响[42]。假设在一个网络中,个体 i 模仿个体 j 的策略:

$$W(s_i \leftarrow s_j) = \frac{w_i}{1+\exp\left[(\beta\pi_{s_i} - \pi_{s_j})\right]}$$

这里,π_{s_i} 和 π_{s_j} 分别表示 i 和 j 的效用,w_i 是 i 受其另一个网络对应个体 i' 影响的因子。w_i 可以由 i' 单独确定,也可以由 i' 所处的局部环境确定。实验结果发现,这种两个网络之间策略信息的分享比传统单一网络互惠能更好地促进系统中合作的涌现。产生这一现象的原因是在相互依赖网络中网络互惠机制与相关联行为自发涌现之间的协同。

Santos 等人则研究了相互依赖网络中允许跨网络策略模仿的有偏模仿规则对合作演化的影响[41]。个体依概率 p 在网络内或概率 $(1-p)$ 网络间选择模仿对象进行策略模仿,并且两个网络上进行的博弈也是不同的,一个是囚徒困境博弈,另一个是雪堆博弈。实验结果显示,随着 p 的增大,两个网络的最终合作水平表现并不相同。在囚徒困境博弈的网络中,即使允许策略在网络传递的概率很小也能够显著地促进该网络中合作的涌现;而在雪堆博弈的网络中,小的 $(1-p)$ 总是会阻碍网络中的合作。

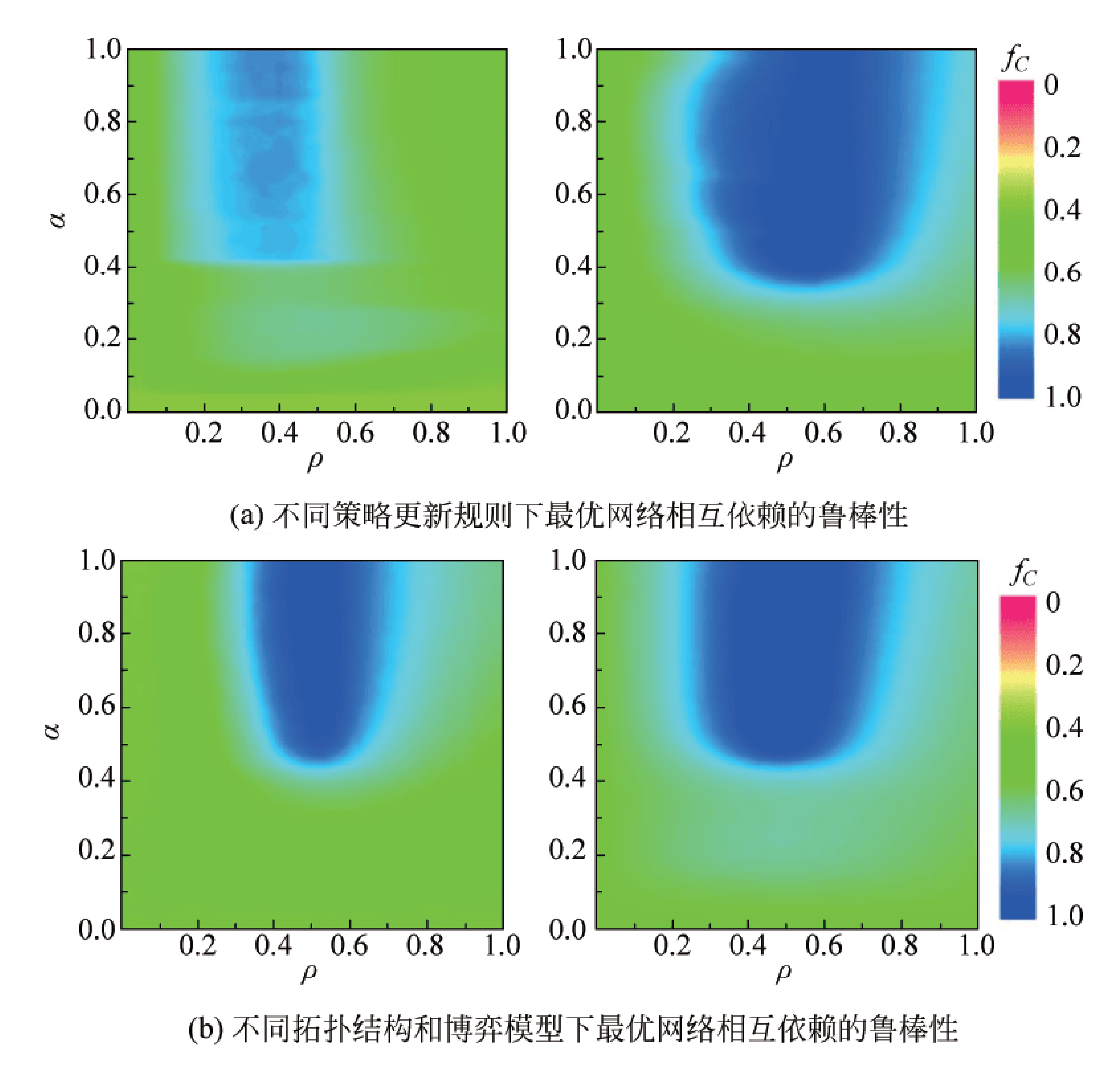


图 3-8　不同条件下最优网络相互依赖的鲁棒性[39]

小结

继第 2 章的经典博弈的相关基础知识之后，本章介绍了演化博弈的相关知识。首先，介绍了演化博弈的基础知识和基本概念，包括演化博弈模型的结构、种群博弈、复制者动态以及演化稳定策略等。然后，介绍了本书的重点内容——网络演化博弈。本章从静态网络上的演化博弈、动态网络上的演化博弈和相互依赖网络上的演化博弈这 3 个层面介绍了网络演化博弈的发展以及近年来的主要工作。

参考文献

[1]　Lewontin RC. Evolution and the theory of games. Journal of Theoretical Biology，1961，1(3)：382-403.
[2]　Smith JM，Price G. The logic of animal conflict. Nature，1973，246：15.
[3]　Taylor PD，Jonker LB. Evolutionary stable strategies and game dynamics. Mathematical Biosciences，1978，40(1)：145-156.
[4]　吴枝喜，荣智海，王文旭. 复杂网络上的博弈. 力学进展，2008，38(6)：794-804.
[5]　Nowak MA. Five rules for the evolution of cooperation. Science，2006，314(5805)：1560-1563.
[6]　Nowak MA，May RM. Evolutionary games and spatial chaos. Nature，1992，359(6398)：826-829.
[7]　Hauert C，Doebeli M. Spatial structure often inhibits the evolution of cooperation in the snow-drift game. Nature，2004，428(6983)：643-646.

[8] Amaral LAN, Scala A, Barthelemy M, *et al*. Classes of small-world networks. Proceedings of the National Academy of Sciences, 2000, 97(21): 11149-11152.
[9] Ohtsuki H, Hauert C, Lieberman E, *et al*. A simple rule for the evolution of cooperation on graphs and social networks. Nature, 2006, 441(7092): 502-505.
[10] Newman ME, Girvan M. Finding and evaluating community structure in networks. Physical Review E, 2004, 69(2): 026113.
[11] Wang Z, Kokubo S, Tanimoto J, *et al*. Insight into the so-called spatial reciprocity. Physical Review E, 2013, 88(4): 042145.
[12] Santos FC, Pacheco JM, Lenaerts T. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. Proceedings of the National Academy of Sciences of the United States of America, 2006, 103(9): 3490-3494.
[13] Chen X, Fu F, Wang L. Prisoner's dilemma on community networks. Physica A: Statistical Mechanics and Its Applications, 2007, 378(2): 512-518. http://dx.doi.org/http://dx.doi.org/10.1016/j.physa.2006.12.024.
[14] Wu J, Hou Y, Jiao L, *et al*. Community structure inhibits cooperation in the spatial prisoner's dilemma. Physica A: Statistical Mechanics and Its Applications, 2014, 412: 169-179.
[15] Wang J, Wu B, Ho DW, *et al*. Evolution of cooperation in multilevel public goods games with community structures. EPL (Europhysics Letters), 2011, 93(5): 58001.
[16] Lozano S, Arenas A, Sánchez A. Mesoscopic structure conditions the emergence of cooperation on social networks. PLoS One, 2008, 3(4): e1892.
[17] Santos FC, Santos MD, Pacheco JM. Social diversity promotes the emergence of cooperation in public goods games. Nature, 2008, 454(7201): 213-216. http://dx.doi.org/10.1038/nature06940.
[18] Shi D-M, Zhuang Y, Wang B-H. Group diversity promotes cooperation in the spatial public goods game. EPL(Europhysics Letters), 2010, 90(5): 58003.
[19] Du W-B, Cao X-B, Hu M-B, *et al*. Asymmetric cost in snowdrift game on scale-free networks. EPL (Europhysics Letters), 2009, 87(6): 60004.
[20] Szolnoki A, Perc M. Coevolution of teaching activity promotes cooperation. New Journal of Physics, 2008, 10(4): 043036.
[21] Zhu C-J, Sun S-W, Wang L, *et al*. Promotion of cooperation due to diversity of players in the spatial public goods game with increasing neighborhood size. Physica A: Statistical Mechanics and Its Applications, 2014, 406: 145-154.
[22] Du W-B, Cao X-B, Zhao L, *et al*. Evolutionary games on scale-free networks with a preferential selection mechanism. Physica A: Statistical Mechanics and Its Applications, 2009, 388(20): 4509-4514.
[23] Wu T, Fu F, Dou P, *et al*. Social influence promotes cooperation in the public goods game. Physica A: Statistical Mechanics and Its Applications, 2014, 413: 86-93.
[24] Wang Z, Zhu X, Arenzon JJ. Cooperation and age structure in spatial games. Physical Review E, 2012, 85(1): 011149.
[25] Pacheco JM, Traulsen A, Nowak MA. Coevolution of strategy and structure in complex networks with dynamical linking. Physical Review Letters, 2006, 97(25): 258103.
[26] Perc M, Szolnoki A. Coevolutionary games—a mini review. BioSystems, 2010, 99(2): 109-125.
[27] Zimmermann MG, Eguíluz VM, San Miguel M. Coevolution of dynamical states and interactions in dynamic networks. Physical Review E, 2004, 69(6): 065102.
[28] van Segbroeck S, Santos FC, Lenaerts T, *et al*. Reacting differently to adverse ties promotes cooperation in social networks. Physical Review Letters, 2009, 102(5): 058105.

[29] Fu F, Wu T, Wang L. Partner switching stabilizes cooperation in coevolutionary prisoner's dilemma. Physical Review E, 2009, 79(3): 036101.

[30] Nowak MA, Sigmund K. Evolution of indirect reciprocity by image scoring. Nature, 1998, 393(6685): 573-577.

[31] Fu F, Hauert C, Nowak MA, *et al*. Reputation-based partner choice promotes cooperation in social networks. Physical Review E, 2008, 78(2): 026117.

[32] Wang X, Wu T, Wang L. Do not aim too high nor too low: moderate expectation-based group formation promotes public cooperation on networks. Physica A: Statistical Mechanics and Its Applications, 2014, 410(0): 259-267. http://dx. doi. org/http://dx. doi. org/10. 1016/j. physa. 2014. 05. 048.

[33] Zhong W, Abbass HA, Bender A, *et al*. Mixed strategy and coevolution dynamics in social networks. Physica A: Statistical Mechanics and Its Applications, 2011, 390(2): 410-417.

[34] Szolnoki A, Perc M, Danku Z. Making new connections towards cooperation in the prisoner's dilemma game. EPL(Europhysics Letters), 2008, 84(5): 50007.

[35] Wu B, Zhou D, Fu F, *et al*. Evolution of cooperation on stochastic dynamical networks. PLoS One, 2010, 5(6): e11187.

[36] Quill E. When networks network: once studied solo, systems display surprising behavior when they interact. Science News, 2012, 182(6): 18-25.

[37] Buldyrev SV, Parshani R, Paul G, *et al*. Catastrophic cascade of failures in interdependent networks. Nature, 2010, 464(7291): 1025-1028.

[38] Wang Z, Szolnoki A, Perc M. Evolution of public cooperation on interdependent networks: the impact of biased utility functions. EPL(Europhysics Letters), 2012, 97(4): 48001.

[39] Wang Z, Szolnoki A, Perc M. Optimal interdependence between networks for the evolution of cooperation. Scientific Reports, 2013, 3.

[40] Wang Z, Szolnoki A, Perc M. Interdependent network reciprocity in evolutionary games. Scientific Reports, 2013, 3.

[41] Santos M, Dorogovtsev S, Mendes J. Biased imitation in coupled evolutionary games in interdependent networks. Scientific Reports, 2014, 4.

[42] Szolnoki A, Perc M. Information sharing promotes prosocial behaviour. New Journal of Physics, 2013, 15(5): 053010.

第 4 章 社交演化博弈

4.1 概述

在社交网络中，用户通过信息分享、信息交互的方式来联系彼此。通常，这些信息都是由用户自身产生和发布的，并能够被其好友接收到。当用户的好友接收到这些信息，并且对这些信息感兴趣时，他们就会回应这些信息，例如对接收的消息进行转发、回复、点赞等。社交网络中的这种信息交互方式反映了用户的社交联系，也满足了他们的社交需求。另一方面，用户在发布和回复消息的时候是需要付出成本的，如时间、费用、精力等。因此，社交网络中用户之间的信息交互行为可以分为两类：合作行为和不合作行为(也称为背叛行为)。

合作广泛地存在于自然界，尤其是人类社会中。通常，合作者需要付出一定的成本为其他人提供帮助，或者相互帮助。而背叛者只是从合作者那里获得收益，但却没有贡献。作为很好的抽象，囚徒困境模型和公共品博弈模型被分别用于研究两两交互和群交互。通常博弈个体有两种策略：合作(Cooperate，C)和背叛(Defect，D)。在博弈中，当两个个体使用的策略都是合作时，他们都能获得奖励 R，而他们都背叛时，则都会得到惩罚 P。当两个个体中一个是合作者，另一个是背叛者时，背叛者通过剥削合作而收获收益 T，而合作者得到 S。当博弈是囚徒困境模型时，$T>P>R>S$。而在公共品博弈中，一组个体同时进行博弈。每个合作者提供的收益被群成员分享，而背叛者只收集收益，不付出任何成本。因此，在这两种博弈模型中，由于背叛者获得的收益比合作者要高，合作者最终会消亡。

演化博弈模型为研究社会困境(social dilemma)[1]中的合作产生和维持提供了通用的模型框架。在演化博弈的研究中，对于不同的情形，研究者们提出了不同的合作演化机制来促进合作的产生，其中包含著名的 5 种促进合作的机制[3]。自从 Nowak 等人[4]提出空间博弈之后，结构种群得到研究者的关注。特别地，图演化博弈(evolutionary graph theory)为研究具有结构的博弈个体的演化机制提出了模型框

架。研究显示,个体之间存在的网络结构在合作的演化过程中扮演着重要角色。进而,不同结构类型的网络得到了研究,如规则网络、随机网络、无标度网络等。人类社会活动形成的网络大多符合无标度网络模型,网络的度分布服从幂律分布[5,6]。

除了网络静态结构的影响,网络的动态机制也吸引了众多研究者的关注。网络结构的动态生长主要是择优连接(preferential attachment)机制[6]。此外,依据个体决策建立、删除或重连个体关系也被广泛研究。Pacheco 等人[7]提出了活跃连接(active linking)机制,在该机制中,个体都有机会随机选择其他个体并按不同的速率与之建立或断开连接。他们的理论分析显示,在连接机制足够快的极端情况下,博弈的收益矩阵发生缩放,这会导致博弈类型发生变化而使得合作者对背叛者占据优势。Fu 等人[8]基于囚徒困境提出了一种共演化模型,该模型允许个体更新他们的策略或者更新他们的合作伙伴。只有合作者被允许单方面地断开已有连接并建立新连接。他们的实验结果显示,连接调整能够使合作维持。在进一步的基于声誉的连接调整的影响的研究[11]中,当个体调整其邻居时,他是基于声誉的情况来断开已有连接和建立新的连接。这里,声誉的定义与 Nowak 和 Sigmund 所给出的"印象分(image scoring)"[10]的定义类似。除了使用囚徒困境模型研究两两交互外,公共品博弈模型还被用来研究网络上的群交互和策略的演化。Zhang 等人[9]基于公共品博弈提出了一种共演化机制。在该机制中,个体评估他们所在群中的产出,并通过适当地调整策略和邻居来获得更多的产出。他们发现,该共演化机制能够传递和促使合作的演化。Li 和 Shen[12]研究了有向网络上的非排他公共品博弈。在他们的模型中,个体根据最大声誉原则将他们的连接关系从"出度邻居(out-neighbor)"转向二步距离的所有邻居。通过与排他性公共品博弈的对比发现,在此连接关系调整机制下,非排他公共博弈表现出阻碍合作的现象。

近年来社交网络服务的繁荣和发展为研究人类社会动力学和社交网络的演化提供了丰富的真实场景[13]。在社交网络服务中,用户可以建立社会关系,相互交流、分享兴趣和信息等。因此,社交网络是一个动态的网络,除了用户的交互活动,他们的连接关系也会发生变化。本章引入了个体策略与网络结构共演化的机制作为社交演化博弈(social evolutionary game,简称 SEG)的模型框架来研究社交网络的动态过程。在社交演化博弈中,每个个体的目的是通过模仿伙伴中更有效率的策略和调整社交关系来分别改进他的短期效用和长期声誉。

4.2 建模方法

本节将详细介绍社交演化博弈模型框架及其基本概念。

在社交演化博弈中,每个个体都有两个关注点,即短期效用和长期声誉。基于个体的关注点,模型中有两种更新机制,即为了获得更好效用的策略更新和更好声誉环境的社交关系更新。

4.2.1 基本概念

定义 4.1(社交演化博弈(social evolutionary game)) 社交演化博弈可表示为在一个带属性的三元组 $G=(V,E,I;U,R)$,其中,$V=\{i \mid 1\leqslant i\leqslant n\}$表示个体集,$E=\{e_{ij} \mid i\in V, j\in V,$

$1 \leqslant i \leqslant n, 1 \leqslant j \leqslant n\}$是个体间的关系(边)集,$I$ 是个体间的交互集,表示个体间的交互依赖与个体间的关系,U 为个体的效用函数集,R 为个体的声誉函数集。

定义 4.2(效用(utility)) 在社交演化博弈中,个体 i 的效用是指在每回合博弈中 i 从对手中获得的收益之和。

效用是个体的短期关注点,它指导个体与其他个体进行策略博弈。个体通过模仿其他成功个体的策略来不断改善自身的效用。

定义 4.3(声誉(reputation)) 在社交演化博弈中,个体的声誉是其他个体对其行为策略历史的印象,通常是对他的合作行为的印象。

声誉是个体的长期关注点,它指导个体更新其社交关系。个体通过与高声誉个体建立社交关系并断开与低声誉个体的社交关系来改善自身所处的社交环境。在 t 时刻,个体 i 的声誉可以形式化为

$$R_i(t) = \sigma R_i(t-1) + \Delta R_i(t)$$

其中,$\Delta R_i(t)$是时刻 t 的声誉增加值,σ 是声誉的记忆衰减率。

4.2.2 博弈设置

假设社交网络中的个体是非理性且自私的,他们在与其他个体进行博弈交互时只关注自身的效用。同时,由于社交网络结构的限制,他们缺乏全局信息而短视,因此只能获取到他们所处环境的局部信息,即他们的邻居节点和邻居节点的邻居节点的信息。

设个体 i 的度表示为d_i。个体有两种交互策略$\{C, D\}$,策略 C 表示合作行为,使用该策略的个体在交互中会付出一定的成本,但是能够使其他个体(或包含自己)获得收益,策略 D 表示非合作行为,使用该策略的个体在交互中不付出任何成本,他只是单纯地获取收益。通常个体之间的交互可分为两类:两两交互和群交互。两两交互也称为成对交互,是指交互发生在两个个体之间,可使用囚徒困境模型、猎鹿博弈、雪堆博弈和协调博弈等进行建模;而群交互是指交互发生在多个个体(超过两个个体)之间,则可以使用公共品博弈进行建模。由此,个体 i 的效用是其参与的所有博弈所获收益之和,即

$$u_i = \sum_{j \in N_i} A_{ij}$$

其中,N_i是 i 参与的所有博弈的集合,A_{ij}是 i 参与博弈 j 所得到的收益。

4.2.3 更新机制

更新机制是异步更新的[11],比如在时刻 t 随机选中一个个体执行更新操作。在社交网络中,个体很难获得全局信息,但通常可以获得其所处的网络的局部信息。因此,基于网络结构和个体获取信息能力的局限性,本文假设个体在进行更新操作时选取的对象个体限制在两步距离内的局部网络环境中。

这里,假设个体的策略更新时间尺度τ_e与社交关系调整的时间尺度τ_p不相同,$W = \tau_e / \tau_p$是策略更新时间尺度和社交关系更新时间尺度之比[14]。当被选中个体进行更新时,他依概率$\frac{1}{1+W}$更新其策略,否则,依概率 $1-\frac{1}{1+W}$调整其社交关系。因此,社交关系更新的频繁

程度随着 W 的增大而增大。此外，依据不同的场景应用设置，在每一演化时步中个体的声誉值会进行相应的更新。

1. 行为策略更新规则

当个体 i 决定更新其策略时，他按照一定的规则模仿其成功邻居 j 的策略。常见的策略模仿规则如下所示。

（1）当社交网络中个体的度是均匀的且不随时间发生变化时，个体 i 依 Fermi 函数确定的概率模仿个体 j 的策略[15]：

$$w(s_i \leftarrow s_j) = \frac{1}{1+\exp[\beta(u_i - u_j)]}$$

其中，β 表示模仿噪声，即个体模仿其他个体的意愿程度。当 $\beta \to 0$ 时，表示趋向随机模仿，而当 $\beta \to \infty$ 时则表示趋向确定性模仿。

（2）当社交网络中个体的度分布是异质的，个体 i 依如下概率模仿个体 j 的策略：

$$w(s_i \leftarrow s_j) = \frac{u_j - u_i}{\mathcal{D}d_{\max}}$$

其中，$d_{\max}$ 是 d_i 和 d_j 中度数大的那个，对于囚徒困境博弈 $\mathcal{D}=T-S$。

（3）当个体 i 更新策略时，他模仿其邻居个体中有着最大效用的最成功个体 j 所使用的策略。不同于以上两种常见的策略模仿规则，该规则是确定性模仿规则。

2. 社交关系更新规则

由于社交网络结构和个体获取信息能力的限制，个体只能获得他们所处环境的局部信息，而非全局信息。因此，在社交演化博弈中，个体获取信息的范围限制在其邻居个体以及邻居个体的邻居个体。

当个体 i 被选中进行社交关系更新时，他依据声誉执行两种操作：建立新的社交关系和断开已有的社交关系。

（1）建立新的社交关系时，个体 i 从其局部环境内未有社交关系的个体中以概率 p_r 按最大声誉原则选择个体 j 作为新的伙伴，否则，以概率 $1-p_r$ 随机选择个体 k 作为新的伙伴；

（2）当断开已有社交关系时，个体 i 从其已有的社交关系中以概率 p_s 按最低声誉原则选择个体 j，并断开与他的社交关系。

4.3　大规模社交行为仿真分析

本节使用无权有向网络表示社交网络用户之间的社交关系，网络中的节点（个体）表示社交网络中的用户。在这个有向网络中，个体 i 有两类社交关系，即 i 的入邻居（IN-neighbors）和出邻居（OUT-neighbors）。如果从个体 j 到个体 i 存在社交关系 e_{ji}，则 j 是 i 的入邻居；反之，i 是 j 的出邻居。N_i^I 表示 i 的入邻居集，N_i^O 表示 i 的出邻居集。d_i^I，d_i^O 分别是 i 的入度和出度，分别表示 i 的入邻居数和出邻居数。d_i 是 i 的度，它是 i 的入度和出度之和。

社交演化博弈是在无权有向网络上进行的，$G=(V,E,I;U,R)$。V 是有向网络中的个体集；E 是个体间的社交关系集；I 表示个体的交互行为集，个体的交互行为与个体所采用

的行为策略有关；U 是个体的效用函数集；R 是个体的声誉函数集。

本节针对社交网络中的两种典型交互场景来说明社交演化博弈模型的建模方法，两两交互和群交互。这里，使用囚徒困境模型描述两两交互，使用公共品博弈模型刻画群交互。

4.3.1　两两交互行为模型设置

网络中的个体与他的邻居依囚徒困境模型进行两两博弈，其效用是他与所有邻居进行博弈所得收益之和，表示如下：

$$u_i = \sum_{j \in \mathcal{N}_i} s_i^T M s_j$$

其中，$\mathcal{N}_i$是 i 的邻居集；s_i和s_j是个体 i 和 j 分别使用的策略；M 是为了方便计算和表示而调整过的 2×2 收益矩阵：

$$M = \begin{pmatrix} 1 & 0 \\ b & 0 \end{pmatrix}$$

当博弈双方相互合作时，他们都获得 $R=1$，而当一方合作、另一方背叛时，背叛者获得 $1 \leqslant b \leqslant 2$。

4.3.2　群交互行为模型设置

网络中的个体与他的邻居依公共品博弈模型进行群交互。个体可以主持一个公共品博弈，即与以他自己为中心以及他的入邻居组成的群体进行公共品博弈，同时也参与其所有的出邻居个体主持的公共品博弈。

当个体 i 主持一个公共品博弈时，所有的个体(包括 i 和他的入邻居)按如下方式获得收益：

$$\begin{cases} p_i^D = r n_i^C b \\ p_i^C = r(n_i^C - 1)b - c \end{cases}$$

其中，p_i^C表示个体采用合作策略 C 时的收益；p_i^D则表示个体采用背叛策略 D 时的收益；n_i^C表示在该公共品博弈中采取合作策略 C 的个体数目；r 表示群体的放大效应因子；c 表示采取合作行为的付出成本；b 则是合作行为提供的利益。依常见设置，$b=c$，$c=1$。

个体 i 的效用是他主持和参与的所有公共品博弈所得收益之和，表示如下：

$$u_i = p_i^{s_i} + \sum_{j \in \mathcal{N}_i^O} p_j^{s_i}$$

在每一演化时刻，按如下方式更新所有个体的声誉：

$$\begin{cases} R_i(t) = \sigma R_i(t-1) + \Delta R_i(t) \\ \Delta R_i(t) = \sin\left(\dfrac{\pi n_i^C}{2 d_i}\right) \end{cases}$$

其中，n_i^C是 i 的邻居中采取合作策略 C 的个体数目，sin()函数是为了压缩声誉的增长速度。

4.3.3　更新机制

策略更新机制和社交关系更新机制的基本规则在 4.2.3 节已给出介绍。这里，针对有向网络的场景进行具体的设置。

1. 策略更新规则

当个体 i 进行策略更新时，他从其出邻居中选择获得最大效用的个体 j，并模仿 j 使用的策略。

$$j = \arg\max_{l \in \mathcal{N}_i^O} \{u_l > u_i\}$$

2. 社交关系更新规则

个体 i 进行社交关系更新时，他依概率 p_r 按最大声誉原则选择个体从其入邻居和出邻居的出邻居中选择未有社交关系的个体 j 并与之建立社交关系。

$$j = \arg\max_{\mathcal{N}_i^I \cup_{l \in \mathcal{N}_i^O} \mathcal{N}_l^O \setminus (\{i\} \cup \mathcal{N}_i^O)} \{R_l(t) > R_i(t)\}$$

否则，依概率 $(1-p_r)$ 随机地从其入邻居和出邻居的出邻居中选择未有社交关系的个体 j' 建立社交关系。同时，i 依概率 p_s 按最小声誉原则从其出邻居中选择个体 k，并与之断开已有的社交关系。

$$j = \arg\max_{l \in \mathcal{N}_i^O} \{R_l(t) < R_i(t)\}$$

4.3.4　仿真结果分析

本节将给出前述示例两两交互和群交互的模型仿真对比结果。初始时，网络设为平均度数 $d=4$ 的随机网络，网络的个体总数为 $n=10^3$，以相同的概率随机分为合作者或背叛者。

当网络的合作者比例 f_c 在一定时间内的波动很小时，则认为网络达到相对稳定状态。否则，演化的仿真时间执行起来相对很长，比如 2×10^5，然后再执行 10^3 次，合作者比例 f_c 取平均值。为了减小随机因素的影响，模型仿真结果是 10 次独立仿真的平均结果。设策略模仿噪声为 $\beta=0.01$，断开社交关系的概率为 $p_s=0.01$，声誉的记忆衰减因子为 $\sigma=1$。

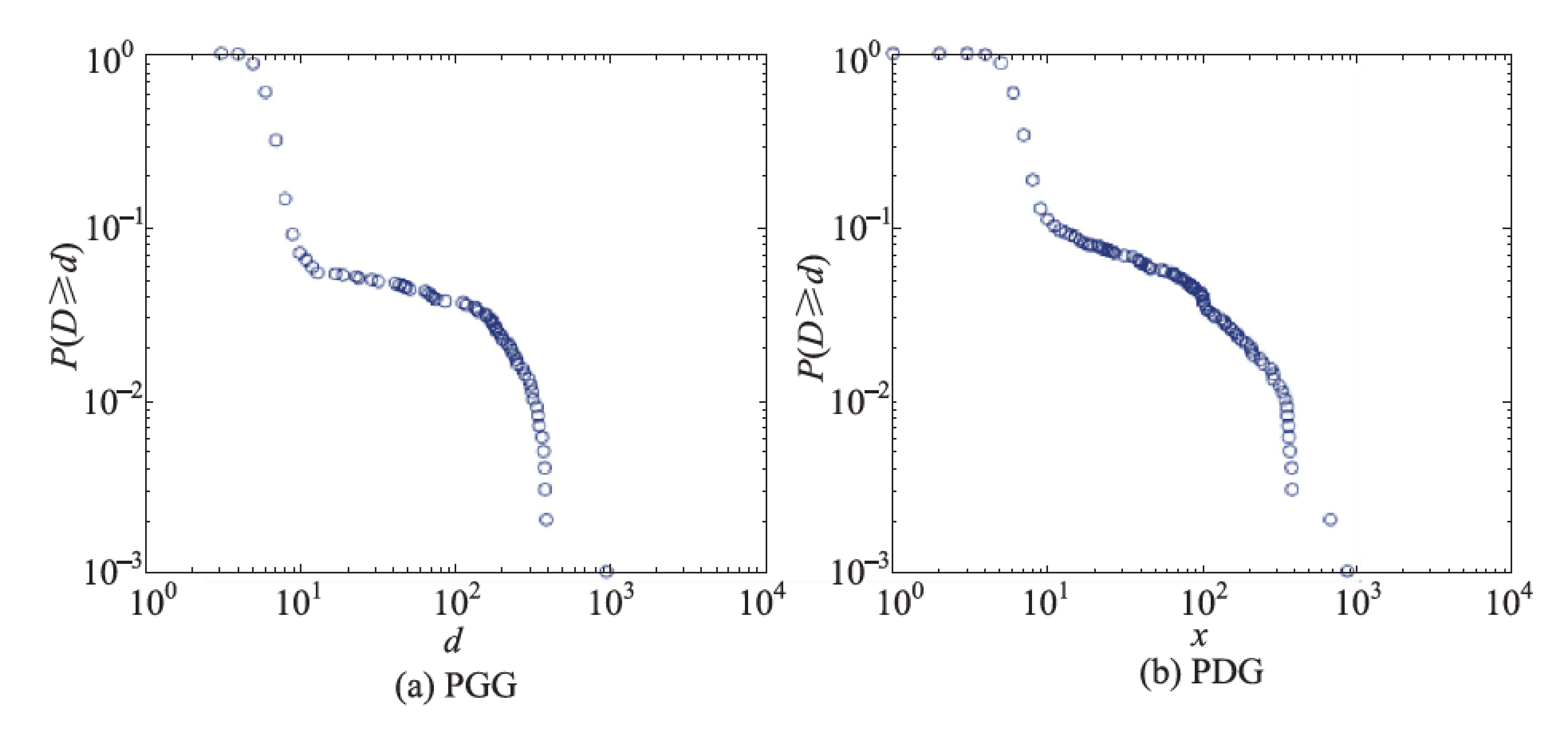


图 4-1　基于 PGG 和 PDG 的演化网络的累积度分布

参数设置：$N=1000$，$W=0.2$，$p_r=0.2$，$r=3.2$ 于公共品博弈，$b=1.2$ 于囚徒困境博弈。采样时间：$t=2\times10^5$

基于社交演化博弈模型的两种更新机制，模型仿真研究了社交网络结构和个体行为策略的共同演化。图 4-1 的两个子图分别显示了囚徒困境模型和公共品博弈模型下网络的累

积度分布。这显示了社交演化博弈模型的个体行为策略与社交关系共同演化的动力机制在一定程度上是可以描述社交网络中的部分场景的演化的。

图 4-2 则显示了这种共同演化的不同时刻的状态截图。为了能够使状态视图表示得更清楚，这里使用较小的个体规模(n=100)。图 4-2(a)～图 4-2(c)显示的是公共品博弈下社交网络演化的状态截图，图 4-2(d)～图 4-2(f)则显示了囚徒困境博弈下社交网络演化的状态截图。从这些子图中可以看到，随着共演化的进行，在公共品博弈模型和囚徒困境模型下网络度数均从同质演化为异质，并都出现了大度数个体。

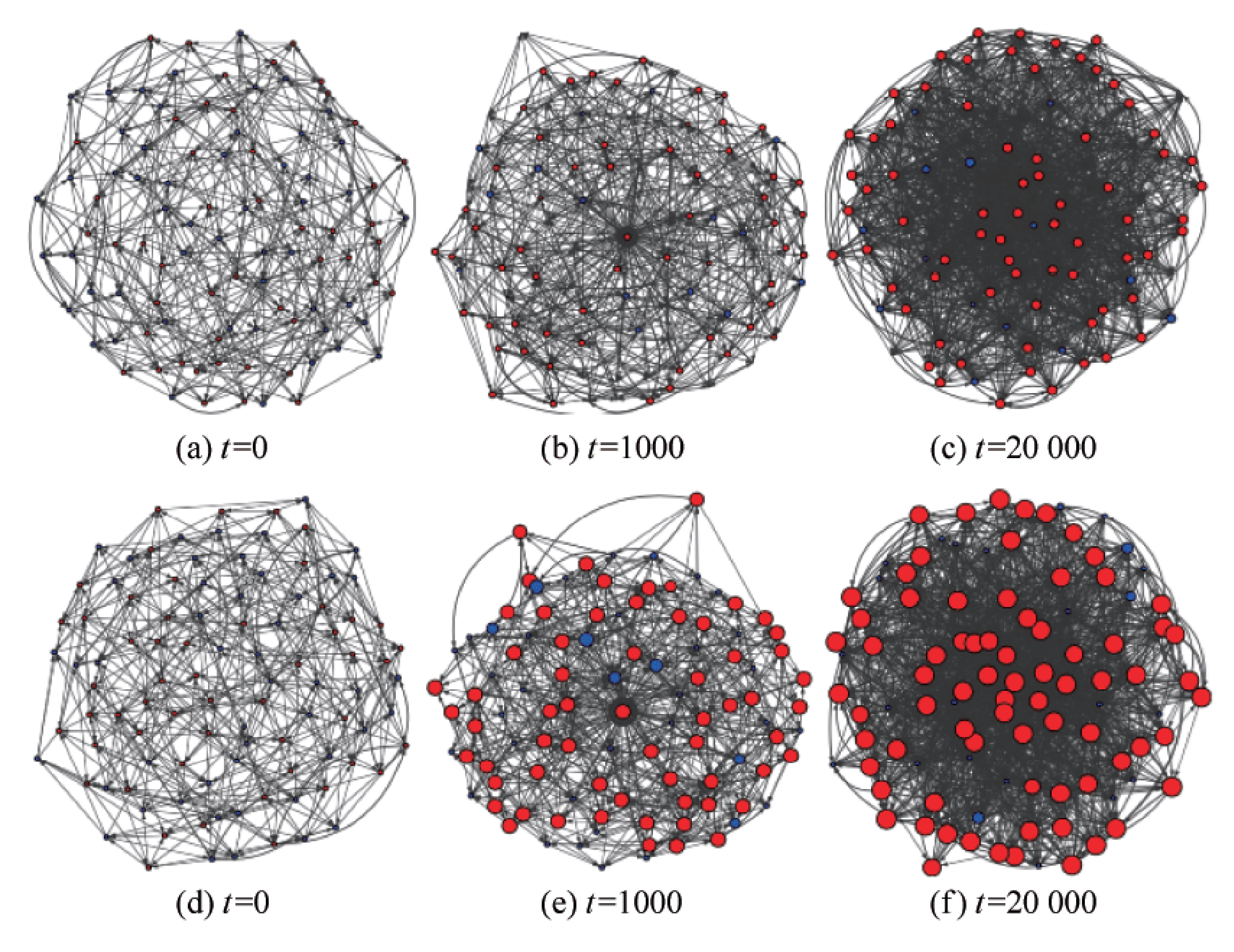


图 4-2　演化过程

上行图是 PGG，下行图是 PDG。合作者标为红色，背叛者标为蓝色。图中个体的大小与个体的声誉值有关，取值是个体声誉值的对数。参数设置：$N=100$，$W=0.8$，$p_r=0.8$，$r=3.2$ 于公共品博弈，$b=1.2$ 于囚徒困境博弈

此外，可以观察到，两两交互和群交互对个体声誉的影响并不相同。两两交互中个体声誉的上升速度明显快于群交互，这意味着在促进个体声誉方面，两两交互比群交互更有效率。

网络中个体合作状态的演化使用合作者比例 f_c 进行评价。在图 4-3(a)中存在一个较小的区域($r<1.8$，$W<0.06$)使得合作难以维持而消失。当 $r>2.6$ 时，即使 $W<0.06$，合作仍然产生。在合作可以生存的区域里，可以观察到网络中的合作水平随 W 的增加而减小。这种现象与非排他性公共品博弈中社交关系更新对合作者比例有阻碍作用相类似[12]。在囚徒困境模型下，社交关系更新也表现出类似的阻碍效果，如图 4-3(c)所示。如该子图所示，当 $b<1.2$ 时，合作者比例随 W 的增加而减小。同时，也存在着合作难以生存的区域($b>1.48$，$W<0.06$)。当 $b>1.48$ 且 $W>0.06$ 时，合作者能够维持，但是相比背叛者并不占据优势。当社交关系更新频率在较低水平时，个体的效用将决定网络中合作者是否能够生存。由于在公共品博弈模型和囚徒困境模型下都出现了对合作的阻碍作用，因此可以认为，这是由于社交关系的有向性导致的，即个体只能调整其出邻居。

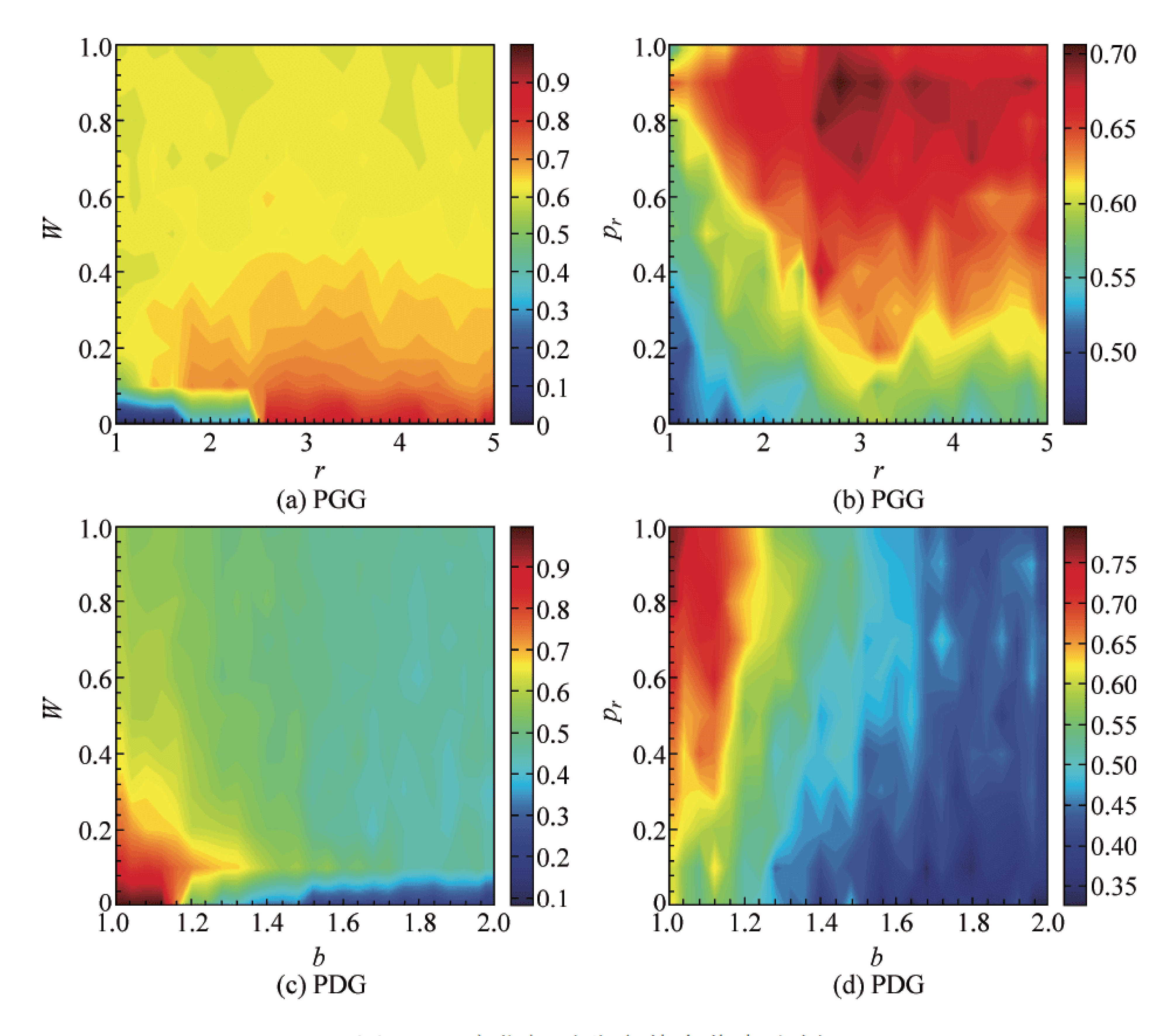


图 4-3　演化相对稳态的合作者比例

PGG：(a)中合作者比例是(r,W)的函数，$p_r=0.8$，(b)中合作者比例是(r,p_r)的函数，$W=0.2$；PDG：(c)中合作者比例是(b,W)的函数，$p_r=0.8$，(d)中合作者是(b,p_r)的函数，$W=0.2$

与 W 对合作的阻碍作用不同，基于声誉社交关系调整的p_r则对合作具有促进作用[11]，如图 4-3(b)和图 4-3(d)所示。声誉作为个体的长期关注点，促使个体在社交网络中产生了更多的交互。公共品博弈模型的群放大效应因子 r 对网络中的合作水平具有促进作用，当 r 增加时，合作者更容易生存，如图 4-3(a)和图 4-3(b)所示。囚徒困境模型的背叛诱惑 b 则对网络中的合作水平有阻碍作用，随着背叛诱惑 b 的增加，合作变得难以维持，如图 4-3(c)和图 4-3(d)所示。群交互对收益的放大效应越大，合作越容易产生。这种现象意味着，如果有恰当的激励机制作用于群交互，则能够促进群体的发展，使网络中的个体变得活跃。相反地，如果在两两交互中，背叛的诱惑越大，合作越难以维持，则网络中的个体就越不活跃。

小结

本章介绍了社交演化博弈模型的基本框架及其在社交网络中两个典型交互场景中的示例。在一个社交演化博弈中，个体只关注他短期的效用和长期的声誉。基于这两个关注点，建立了个体行为策略更新和社交关系更新的共演化机制。对比了社交演化博弈在两种典型的社交网络交互场景下的建模与仿真，即两两交互和群交互。囚徒困境博弈模型被用来建模两两交互，而公共品博弈模型则用来描述群交互。通过对两种博弈模型的社交演化博弈

的模型仿真，可以观察到社交演化博弈的更新机制会导致社交网络的累积度分布呈近似幂律分布。同时，社交关系调整对网络中的合作水平有阻碍效应，而对声誉的追逐则会促进合作的繁荣。此外，在群交互中随着群放大效应的增大，合作更容易生存，而在两两交互中背叛的诱惑越大，合作越难以维持。

参考文献

[1] Dawes RM. Social dilemmas. Annual Review of Psychology，1980，31(1)：169-193.

[2] Kollock P. Social dilemmas：the anatomy of cooperation. Annual Review of Sociology，1998：183-214.

[3] Nowak MA. Five rules for the evolution of cooperation. Science，2006，314(5805)：1560-1563.

[4] Nowak MA，May RM. Evolutionary games and spatial chaos. Nature，1992，359(6398)：826-829.

[5] Barabási A-L，Albert R. Emergence of scaling in random networks. Science，1999，286(5439)：509-512.

[6] Albert R，Barabási A-L. Statistical mechanics of complex networks. Reviews of Modern Physics，2002，74(1)：47.

[7] Pacheco JM，Traulsen A，Nowak MA. Coevolution of strategy and structure in complex networks with dynamical linking. Physical Review Letters，2006，97(25)：258103.

[8] Fu F，Wu T，Wang L. Partner switching stabilizes cooperation in coevolutionary prisoner's dilemma. Physical Review E，2009，79(3)：036101.

[9] Zhang C，Zhang J，Xie G，*et al*. Coevolving agent strategies and network topology for the public goods games. The European Physical Journal B，2011，80(2)：217-222.

[10] Nowak MA，Sigmund K. Evolution of indirect reciprocity by image scoring. Nature，1998，393(6685)：573-577.

[11] Fu F，Hauert C，Nowak MA，*et al*. Reputation-based partner choice promotes cooperation in social networks. Physical Review E，2008，78(2)：026117.

[12] Li Y，Shen B. The coevolution of partner switching and strategy updating in non-excludable public goods game. Physica A：Statistical Mechanics and Its Applications，2013，392(20)：4956-4965.

[13] Leskovec J，Backstrom L，Kumar R，*et al*. Microscopic evolution of social networks. In：Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. 2008：462-470.

[14] Santos FC，Pacheco JM，Lenaerts T. Cooperation prevails when individuals adjust their social ties. PLoS Computational Biology，2006，2(10)：e140.

[15] Szabó G，Töke C. Evolutionary prisoner's dilemma game on a square lattice. Physical Review E，1998，58(1)：69.

第二部分

单一网络的社交网络演化计算

第一部分回顾了社交网络及其演化、博弈和演化博弈等的相关知识，提出了面向社交网络演化研究的社交演化博弈模型。本部分在第一部分的基础上讨论单一网络的社交网络演化计算模型、方法和实施用例，包括社交网络中信息传播的演化计算、信息行为与结构的演化计算、群体以及重叠群体的演化计算。

第 5 章介绍社交网络信息传播的演化计算模型、方法和实施用例。首先介绍社交网络的典型拓扑结构及拓扑结构对于信息传播的影响，进而提出基于网络边传播能力的异质性的传播模型，即信息传播的一致性模型。该模型通过对用户注意力、网络边的传播能力以及传播能力的异质性进行建模。随后介绍了竞争信息传播模型(ICSIS)及该模型的稳定性分析，进一步讨论信息发布时间点和网络节点特征对于竞争性信息传播的影响。最后利用社交演化模型对竞争性信息异步传播和同步传播两种传播形式的演化过程进行了分析。

进一步地，在第 6 章讨论社交网络中信息行为与网络结构的共演化。首先讨论社交网络用户典型信息行为的分类，而后讨论了不同因素影响下的用户关系更新方式，进而利用社交演化博弈分别给出了信息分享行为、信息交互行为与社交网络结构演化的相互影响分析。

更进一步地，第 7 章给出社交网络群体演化的计算方法，首先从群体结构和群体行为两个层面给出社交网络群体的评价指标，进而给出基于博弈的信息交互群体识别方法，该方法利用用户信息交互行为自底而上地发现信息交互紧密的群体，并在 Facebook 数据集上进行验证。进而基于网络重叠群体，讨论重叠群体相互影响的方式，给出重叠群体相互影响演化的计算模型，分析重叠群体的共演化现象。最后讨论跨网络用户关注度竞争的问题。基于社交演化博弈，提出一个竞争性社交化博弈模型，用以研究多个社交网络之间对于用户关注度的竞争过程。同时给出用户关注度的概念，以及社交网络的流行度概念，用以描述社交网络用户在社交网络服务上的时间分布。

第 5 章 网络结构对信息传播的影响

社交网络上的信息传播在诸多因素的影响下进行着演化，既包括传播速度和范围的演化，也包括信息自身内容的演化。影响因素有很多，但归结起来不外乎信息自身的特征、传播信息的网络用户的特征与行为、承载信息传播的社会网络的拓扑结构，以及信息传播的宏观环境。另外，社交媒体中，信息传播还尤其受到社会媒体服务所提供的信息推送功能（如 Facebook 的 News Feed）的影响。这几个方面是社交网络中信息传播的关键因素，它们共同决定了信息传播与演化的行为与模式。另一方面，在社交网络中，有无数信息同时在传播。这些信息之间并不完全都是相互独立的，相反，有些信息之间存在着关联关系。这种关联关系可以分为正、负两大类。所谓正关联关系，是指那些相互之间能够形成促进作用的关系；而负关联关系则是指彼此之间存在着竞争或抑制的关系。信息间的关联关系对它们的传播演化行为与模式是很有影响的。然而，现有研究信息传播的工作，基本上都仅仅研究单一信息的传播模式，不能将其通过简单的扩展来研究关联信息的传播模式，而关联信息的传播比单一信息传播有着更为真实的应用背景。此外，在社交网络中，一方面，同一时间内从特定主体产生的相关信息同时在网络中传播，用户与其网络邻居针对这些竞争信息进行交流和分享，形成同步信息竞争传播；另一方面，某一时间段内从特定主体产生的相关信息，先后在网络中传播并相互竞争，用户针对此类竞争信息进行交流和分享，形成异步信息竞争传播。本章特别关注社交网络拓扑结构对信息传播演化的影响，同时研究竞争性信息的传播演化机制。

5.1 信息传播与网络结构

由于现实场景的复杂性，研究者们提出了具有不同结构的网络模型及其生成方式。规则网络、随机网络、小世界网络以及 Scale-Free 网络等都具有十分典型的拓扑结构。其中，规则网络的边连接比较规则，

一般文献提到的规则网络其平均聚集系数较大,平均路径较长;随机网络的边连接具有随机性,其聚集系数较小,平均路径较短;小世界网络呈现出明显的社团结构,其聚集系数较大,而平均路径较短。Sacle-Free 网络结构则体现出其规模(scale)上的特点。给定一个 Scale-Free 网络,无论该网络的节点如何增加,其节点的度分布均是稳定不变的,所以称其为 Scale-Free 网络。

在正式介绍上述几种网络之前,首先给出度、度的分布、平均聚集系数和平均路径的定义以及计算方式。

度与度分布　假设给定一个含有 N 个节点的网络 G(本节介绍的所有网络均指的是无向图网络),任意一个节点 i 的度表示为k_i,在数值上等于与该点连接的边的数量。度的分布 $P(k)$指的是,从图中随机挑选一个节点,其度为 k 的概率。

平均路径　假设给定一个含有 N 个节点的网络G,定义 d_{ij}为任意两点 i 和 j 之间的距离,在数值上等于该两点之间的最短路径的长度。那么,网络 G 的平均路径 L 可以通过如下公式计算:

$$L = \frac{2}{N(N-1)} \sum_{i<j} d_{ij}$$

平均聚集系数　同样,假设给定一个含有 N 个节点的网络 G,对于任意一个节点 i,它的度表示为 k_i,这代表它有k_i个邻居。在这k_i个邻居之间,理论上最多存在 $k_i(k_i-1)/2$ 条边。令E_i表示这 k_i个邻居之间实际的边的数量,则实际值与理论最大值的比值,就代表了该点的聚集系数。即点 i 的聚集系数计算公式如下:

$$C_i = \frac{2E_i}{k_i(k_i-1)}$$

网络的平均聚集系数则是所有点的聚集系数的算术平均值,即

$$C = \frac{\sum_{i=1}^{N} C_i}{N}$$

5.1.1　典型网络拓扑结构

1. 规则网络

规则网络是指规则连接的网络,常见的规则网络可分为以下 3 种:

(1) 全连接网络　在该网络中,每个节点都与其他所有节点连接,所以该种网络具有最大的聚集系数以及最短的平均最短路径,其值均为 1。

(2) 环状网络　如果每个节点均和最近的 k 个节点连接,则组成了一个规则的环状网络。

(3) 星状网络　在该网络中,有一个中心节点,除该节点外,其他所有节点均只和该中心节点连接。

2. 随机网络

与规则网络完全不同的是,随机网络的边连接具有完全的随机性。ER 随机网络便是一种典型的随机网络[1],可通过如下步骤产生:

（1）首先产生 N 个孤立的节点；

（2）N 个节点最大可以产生 $N(N-1)/2$ 条候选边。以概率 $p(0\leqslant p\leqslant 1)$ 去连接所有的候选边。

如此产生的随机网络具有 $pN(N-1)/2$ 条边。注意，当 $p=0$ 时，生成的网络仍然是 N 个独立的节点；而当 $p=1$ 时，是一个完全连接的规则网络。

3. 小世界网络

规则网络具有较大的聚集系数，而随机网络具有较短的平均路径。小世界网络可以同时具有这两种拓扑特征。这种网络首次由 Watts 和 Strogatz 在 1998 年提出[2]。

4. WS 小世界网络

WS 小世界网络可以通过如下步骤产生：

（1）构造一个环状的规则网络；

（2）对于上述规则网络中的每一条边（a-b），均以概率 p 重新连接到另外一个随机挑选的节点（c），即打破原有边（a-b），重新连接成边（a-c）。

当 $p=0$ 时，该网络仍然是规则网络；当 $p=1$ 时，该网络和 5.1.1.2 节中的随机网络具有相似的拓扑结构；当 $p\in(0.01,0.1)$时，该网络具有很好的小世界特征：较大的平均聚集系数和较小的平均路径。

5. NW 小世界网络

在 WS 小世界网络生成的过程中，网络的联通性可能会被破坏（边的重连可能会导致多个独立的网络）。于是，Newman 和 Watts[3] 对 WS 小世界网络的生成算法做出了细微的改变，并称其为 NW 小世界网络。

（1）构造一个环状的规则网络，该网络中的 K 值一般较小；

（2）对于网络中所有没有连接的点对，以概率 $p(0<p\ll 1)$去连接这些点对。

注意，当 $p=0$ 时，产生的网络仍然是环状规则网络；当 $p=1$ 时，产生的是完全连接的规则网络；当 p 值很小时，该算法产生的网络性质与 WS 小世界网络几乎一致。

6. Scale-Free 网络

上面提到的随机网络与小世界网络，其节点度的分布均是泊松分布。度分布的最大值出现在 $k=\bar{k}$（平均度）的时候。具有很高度（$k\gg\bar{k}$）的节点几乎是不存在的。但是，现实中的很多网络，其度的分布往往是幂分布，并且这种分布的数值特征与网络的大小没有关系，所以称其为 Scale-Free 网络。

不仅如此，随机网络与小世界网络的大小不会随着时间而改变。但是，在万维网等网络中，节点是会随着时间而增加的。在网络的增长过程中，新加入的节点更倾向于去连接度数较高的节点。

文献[4]首次提出了描述这种现象的网络。这种网络被称为 BA Scale-Free 网络，其产生过程如下：

（1）构造一个具有 m_0 个节点的完全连接规则网络。每次均加入一个新的节点，该节点

会从当前的网络中选择 m 个节点连接($1\leqslant m\leqslant m_0$)。

(2) 在步骤(1)描述的新节点去连接已有的 m 个节点过程中,对于网络中已有的节点 i 和它的度数 k_i,它被新节点连接的概率是

$$p=\frac{k_i}{\sum_{j=1}^{N}k_j}$$

无论加入多少新的节点,该算法产生的网络的度分布均是固定的幂分布

$$P(k)\sim 2m^2k^{-3}$$

5.1.2 拓扑结构对信息传播的影响

网络拓扑结构对信息传播模式的影响是信息传播的一个研究热点。目前已有大量工作研究了不同的网络拓扑结构对信息传播的影响[5-8],譬如无标度网络[7-9]、小世界网络[10]、随机聚团网络[11]、层次结构网络[12]、社区结构网络[13]以及其他类型的网络,如具有非随机远程互动的网络[14]等。Huang 等人研究了信息在具有社区(模块)结构的网络中传播的动力学特性[13]。北京大学的 Fu、Liu 和 Wang 研究了层次状网络中的信息传播[15]。在这种网络中,他们发现,初始所选定的感染种子的位置对信息扩散的范围有着很大的影响。当初始选定的种子属于某个过渡层时,信息能够达到最大的扩散范围。国防科学技术大学的 Fan 和 Li 等人研究了新浪微博[16]。他们发现,新浪微博网络具有明显的小世界效应和无标度特征,而且出度分布似乎有多个独立的具有不同指数的幂律状态。Volchenkov 等人发现,在无标度网络中流行病的扩散对节点度的分布是非常敏感的[9]。在文献[10]中,电子科技大学的 Lü 和 Zhou 等学者发现,当网络的尺度逐步增大时,随机网络相比规则网络更能促进信息的有效扩散。信息扩散的有效性能够通过在规则结构中引入些许随机性而被急剧加强。因此,他们认为小世界网络导致了最有效的信息扩散。而在文献[17]中,Karsai 和 Barabasi 等人发现,尽管小世界结构网络具有较短的路径,但在这种网络中,信息扩散反而减缓了。他们发现,在小世界网络中,信息传播的减缓主要由权重与拓扑结构的相互关系以及个体突发式的行为模式所引起。上面诸多的研究工作都仅仅考虑了不同的拓扑结构对信息传播的影响作用,但毫无疑问,信息自身的特征以及传播信息的网络用户的特征与行为等因素对信息传播也起着至关重要的作用,而传统的工作几乎都没有考虑到这一点。

5.1.3 信息传播的一致性模型

1. 模型综述

Ou 与 Jin 等人[18]提出了基于网络边传播能力的异质性的传播模型(下面所指的一致性模型均是指该篇文献提出的模型)。他们认为,异质性的强弱导致可观察到不同的传播行为特征。这种异质性是指网络中不同的边具有不同的传播能力。当异质性较弱时,信息传播在小世界网络更加有效;而当异质性较强时,信息传播在随机网络更加有效。

该模型主要通过对用户注意力、网络边的传播能力以及传播能力的异质性进行建模。

该模型用到的符号如表 5-1 所示。

表 5-1　符号说明

符号	含义
Q_{max}	节点最大的注意力
Q_i	节点当前被激活的注意力
$node_i$	一个节点 i
q_{ij}	从节点 j 到 i 的传播能力
f_{ij}	节点 i 与 j 之间的共同好友数
n_i	节点 i 的好友数

2. 用户注意力

文献[19]在 Twitter 网络上进行了一个统计研究，作者发现，每个用户对信息的注意力（或者说处理信息的能力）都是一样的，且大致相等。在模拟实验中，作者允许每个用户去竞争其邻居的注意力，即每个用户都有一定的概率去转发（且只转发）某一个邻居发送过来的信息。与该文献的做法类似，在一致性模型中，作者假设网络中的每一个点均具有一个统一的最大注意力，记作$Q_{max}=1$。每个点的任意一个邻居，都有能力去唤醒该点一定数量的注意力。假设有一对邻居$node_i$与$node_j$。如果$node_j$传播了一条信息，那么$node_i$就会注意到该条信息，$node_i$大小为q_{ij}的注意力已经被激活，即$node_i$表示该点对该条消息的注意程度。显然，q_{ij}就代表了该条边的传播能力（从$node_j$到$node_i$），这种能力与这对邻居的关系呈正相关，具体如何量化将在下面第 3 点进行建模说明。

3. 对边传播能力的异质性的建模

接下来，从一个点$node_i$以及它的n_i个邻居（$node_{j_1}$，$node_{j_2}$，…，$node_{j_{n_i}}$）开始建模。给定一个网络，每条边应该具有不同的传播能力，该能力记作 q_{ij}。具体量化每条边的传播是困难的，但在网络中，关系亲密的点经常组成一个社区。信息在这种社区内很容易传播，在社区之间则传播困难[20]。文献[21]指出，两个点之间重叠的邻居个数（即共同好友）可以用来衡量两个点之间的亲密程度。具体地，关系紧密的两个点的传播能力要大于关系疏远的点对之间的传播能力。因此，可以让 q_{ij} 与两个点$node_i$，$node_j$之间的共同好友数成正比例关系，即

$$q_{ij_k}=\frac{f_{ij_k}}{\sum_{r=1}^{n_i} f_{ij_r}} \tag{5-1}$$

然而，直接使用该公式有一个缺陷。假设两点之间的共同好友数为 0，即$f_{ij}=0$，这种情况在随机网络中是十分常见的。那么在这种情况下，这条边的传播能力就下降为 0，信息就无法在这条边上传播。所以要尽量避免这种情况的发生，即使是关系疏远的点，只要有边连接，那么这条边也有较小的传播能力。于是，可采用一个参数 $\beta(\beta>0)$来平滑这个公式，即

$$q_{ij_k}=\frac{\frac{1}{\beta}+f_{ij_k}}{\frac{1}{\beta}\times n_i+\sum_{r=1}^{n_i} f_{ij_r}} \tag{5-2}$$

其中，n_i表示$node_i$的好友总数。异质性体现在对于不同的边，其传播能力不同，通过q_{ij}量化。而参数β则可以调节控制这种异质性。如果β较大，则q_{ij}对于f_{ij}比较敏感，即不同的f_{ij}值会造成差异很大的q_{ij}值。而当β较小时，q_{ij}对f_{ij}不敏感，在这种情况下，任何点的所有边的传播能力都是一样的，其值约等于$\frac{1}{n_i}$。所以，当β较大时，代表异质性很强；当β较小时，代表异质性很弱。该公式还有一个好处，对于任何一点，无论β如何变化，$\sum_{r=1}^{n_i} q_{ij_r}=1$始终成立。这也就是说，每个点对于信息的处理能力都是有限且相等的。

4. 对传播概率的建模

用户对于同条信息以往的接触，都会增加该用户转发该信息的概率[22]。其他研究表明，社会强化因素也会加大用户转发信息的概率[10,23,24]。因此，这些同条信息之前的影响体现在Q_i上。该值是指对于每个节点，已经积累的被激活的注意力的总和。所以，其计算公式如下：

$$Q_i=\sum_k q_{ij_k} \tag{5-3}$$

式中，k是指所有已经传播了信息的邻居$node_{j_k}$。上式保证了对于任何一个用户，Q_i的上限均为1。同时，提出下面的概率转化公式

$$P_i=\tanh(\alpha\times Q_i^2) \tag{5-4}$$

在该公式中，Q_i与P_i是平方关系，而不是线性关系，这是一个经验性的建模。也就是说，很小的Q_i只能造成很小的传播概率，而当Q_i很大时，才能造成很大的概率。参数α代表信息本身的流行程度，是信息的固有属性。如果一个信息本身很流行，那么在Q_i不变的情况下，其被传播的概率更大，即更流行的信息被传播的概率更大[25]。实验重点考察异质性(即β的取值)。即使在不同的α取值下，模型捕获的传播行为是一致的，除非α被设计得过大或者过小。tan$h$函数保证了概率的取值始终在0与1之间。

上面的模型有很多有点。首先，参数β可以控制异质性的强弱。传播概率函数是关于注意力大小的一个单调递增函数，如果用户被激活的注意力更多，其传播该条信息的概率更大。tan$h$函数保证概率的取值在0～1之间。

5. 实验设计

本节简单叙述该模型中的参数是如何选择的、实验所用的网络是如何生产的，以及模拟规则。

- 参数选择

在实验中，α的取值范围是6～9，β的取值范围是0.1～1。在这些值的区间范围内，该模型可以观察到不同的传播行为。虽然该模型与经典的传播模型适用的场景不一样，如SIR模型，但是仍然可以借用传播理论上的阈值理论来论证参数设计的合理性[26,27]。在文献[26,27]中，作者提出传播阈值约等于网络平均度的倒数。在下面将要用到的网络和文献[10]中，网络平均度记为K，其值为6。当传播的概率大于$\frac{1}{K}$时，信息或病毒会传播到网络几乎所有的点；而当传播概率小于$\frac{1}{K}$时，信息或病毒就很快消失，只会传播到网络上很小

的一部分区域。本节中,α 的取值是 6～9。假设当 β 接近于 0 的时候,当一个点第 1 次接收到一条信息时,其传播的概率为 $P=\tanh\left(\frac{\alpha}{K^2}\right)$。如果 $\alpha=6$,则有 $P=\tanh\left(\frac{1}{6}\right)\approx 0.1651$,该值十分接近于$\frac{1}{K}$。所以 α 的设计在 6～9 之间。当 α 大于 9 时,其传播就会覆盖到网络中几乎所有的点,在这种情况下,该模型无法捕捉到异质性对传播的影响。其次,β 的取值在 0.1～1 之间,则$\frac{1}{\beta}$的取值范围为 10～1 之间,与 $K=6$ 相比,相信该值可以很好地控制异质性的强弱。

- **模拟图产生算法**

实验所用的网络图通过文献[2]中的算法实现。首先,构建一个节点数量为 N 的规则网络,即每个节点与其最近的 K 个节点相连($K=6$, $N=10\,000$)。然后对该规则网络中的每一条边执行下面的步骤:对于每一条边(a-b),将该边以概率 ρ 进行重连,即以概率 ρ 决定是否对该边进行重连。如果要重连,则随机选择点 c,打破原来的边 a-b,连接成 a-c(如果 a-c 已经存在,则放弃这一步)。ρ 的值就决定了最后生成网络的类型。ρ 越大,生成的网络更加具有随机性。当 $\rho\in(0,0.01)$时,生成的网络是规则网络;当 $\rho\in[0.01,0.1]$时,生成的是小世界网络;当 $\rho\in(0.1,1]$时,生成的是随机网络[2,28]。

- **模拟规则**

接下来将描述模拟过程。当一个节点传播一条信息的时候,它的邻居都能看见这条信息,然后通过模型去判断是否转发。如果该点也转发了这条信息,那么它的邻居也都能看到这条信息。如此,该条信息就传播下去,直到没有新的节点去传播这条信息。已经传播了信息的节点,标记为死节点,不会参与到接下来的模拟过程。具体而言,首先,所有节点的Q_i在初始状态均设置为 0。随机选择一个节点,作为种子节点。该节点会把信息传播给它的邻居。对于接下来的每一步,在上一步收到信息的节点会更新自己的已经激活的注意力总量 Q_i(公式(5-2)、公式(5-3)),然后,根据公式(5-4)产生一个概率来决定自己是否转发。当没有新的节点传播信息的时候,模拟过程结束。使用网络中传播了信息的节点总数来表示哪种网络更适合信息传播(即比较不同网络上,传播信息的节点总数)。

6. 实验与结论

本节将简述模拟网络上的实验,并讨论实验的结果。对于生成的不同类型的网络(以 ρ 值区分),对每一个 α 与 β 的组合,均进行 1000 次独立实验。下面章节展示的结果,均是这 1000 次实验的均值。

(1) 哪种网络更适合信息传播

首先研究在不同异质性强弱的影响下,哪种网络更适合信息传播。显然,当异质性很强的时候,如图 5-1(a)、图 5-1(b)($\beta=0.9,0.7$)所示,无论 α 的取值如何变化,曲线的最高点均出现在随机网络上($\rho\in(0.1,1]$)。这表明,在异质性比较强的时候,随机网络比小世界网络更适合信息传播。

而当异质性很弱的时候,如图 5-1(c)、图 5-1(d)所示,信息在小世界网络上传播的范围更广($\rho\in[0.01,0.1]$),这与文献[10,23]的结论是一致的。

参数 α 可以控制总体的传播范围。当 α 很小时,传播的范围很窄;而当 α 取值很大时,

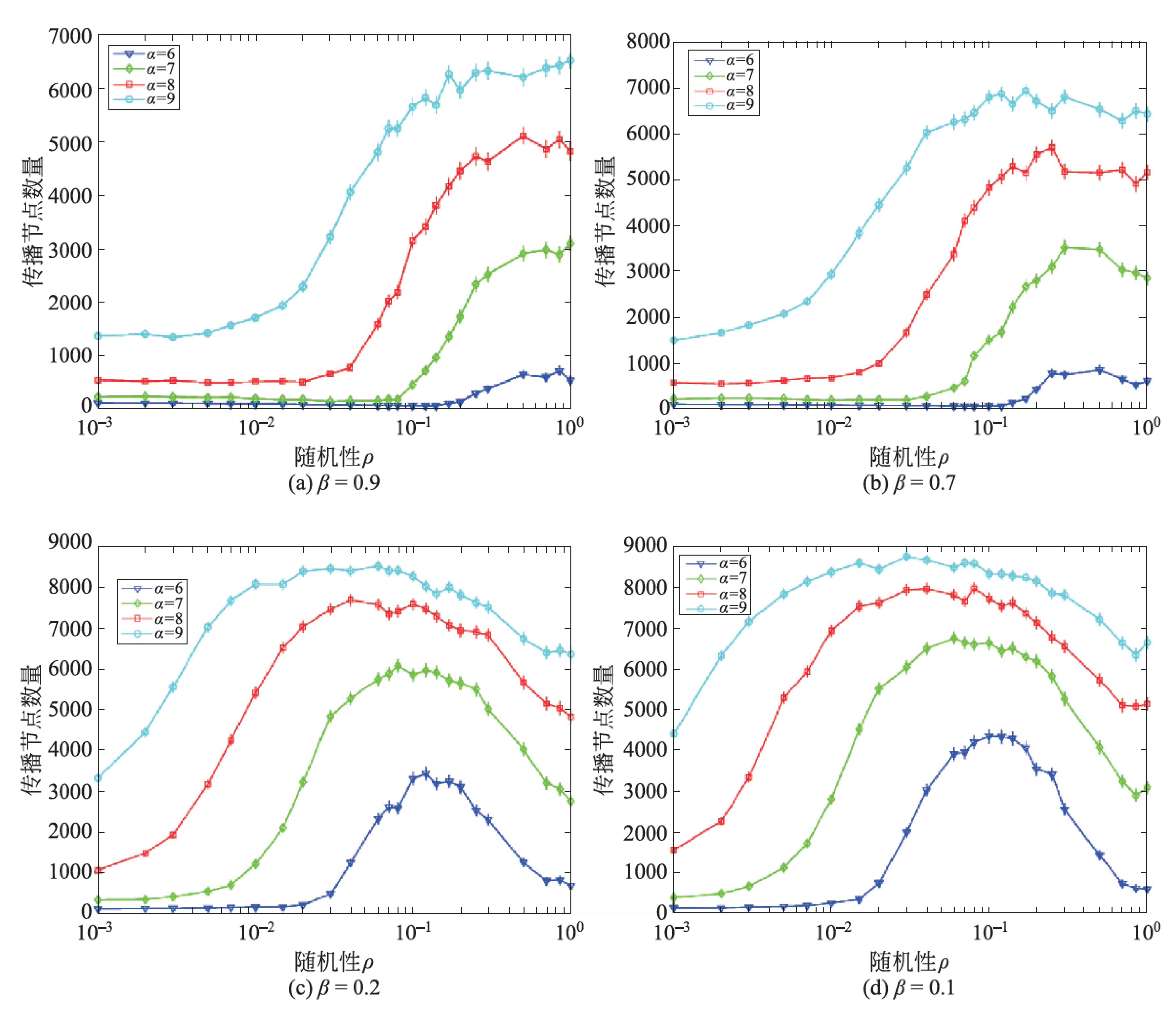


图 5-1　传播节点数量与随机性 ρ 的关系。其他参数是 $K=6$，$N=10\,000$。实验结果取 1000 次独立实验的平均值

其传播的范围很广。根据上节参数设计部分的讨论，第 1 次传播的概率不能设计得太大或太小，否则，信息会传播到所有的节点或者无法传播。例如，如图 5-1(d)所示，当 $\alpha=9$ 时，即使在规则网络上，信息也传播到了 40%以上的节点。

(2) 异质性如何影响信息传播

接下来，通过固定 α，调节 β，来研究异质性是如何影响信息在不同类型的网络上传播的。

如图 5-2 所示，当 α 值确定后，随着异质性的加强(β 从 0.1～0.9)，小世界网络与规则网络的传播越来越受到抑制。这表明，对于小世界网络而言，异质性与传播效率是负相关的，异质性越强，小世界网络的传播越受到限制。

对于随机网络而言(ρ 接近 1 的时候)，信息在该类网络上的传播并没有受到影响。这表明，异质性的强弱对于随机网络并没有明显的影响。

(3) 异质性对传播速度以及传播周期的影响

接下来，从传播速度以及信息的生命周期两个角度来研究异质性对信息传播的影响。

上一节介绍了该实验的模拟方法。定义消息的生命周期为每次模拟的步数均值，而信息的传播速度为每一步传播该信息的节点数量的均值。

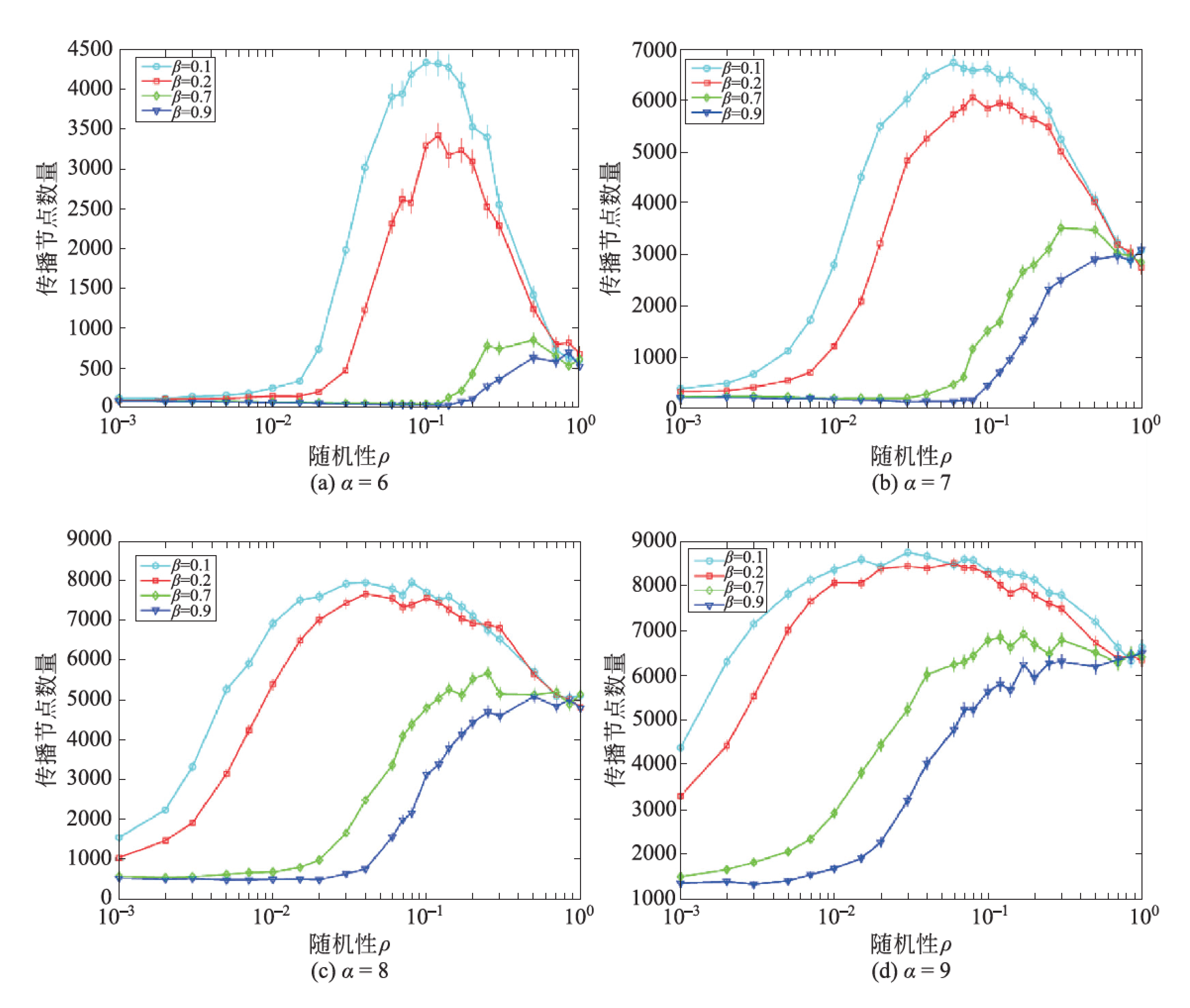


图 5-2　传播节点数量与随机性 ρ 的关系。其他参数是 $K=6$，$N=10\,000$。实验结果取 1000 次独立实验的平均值

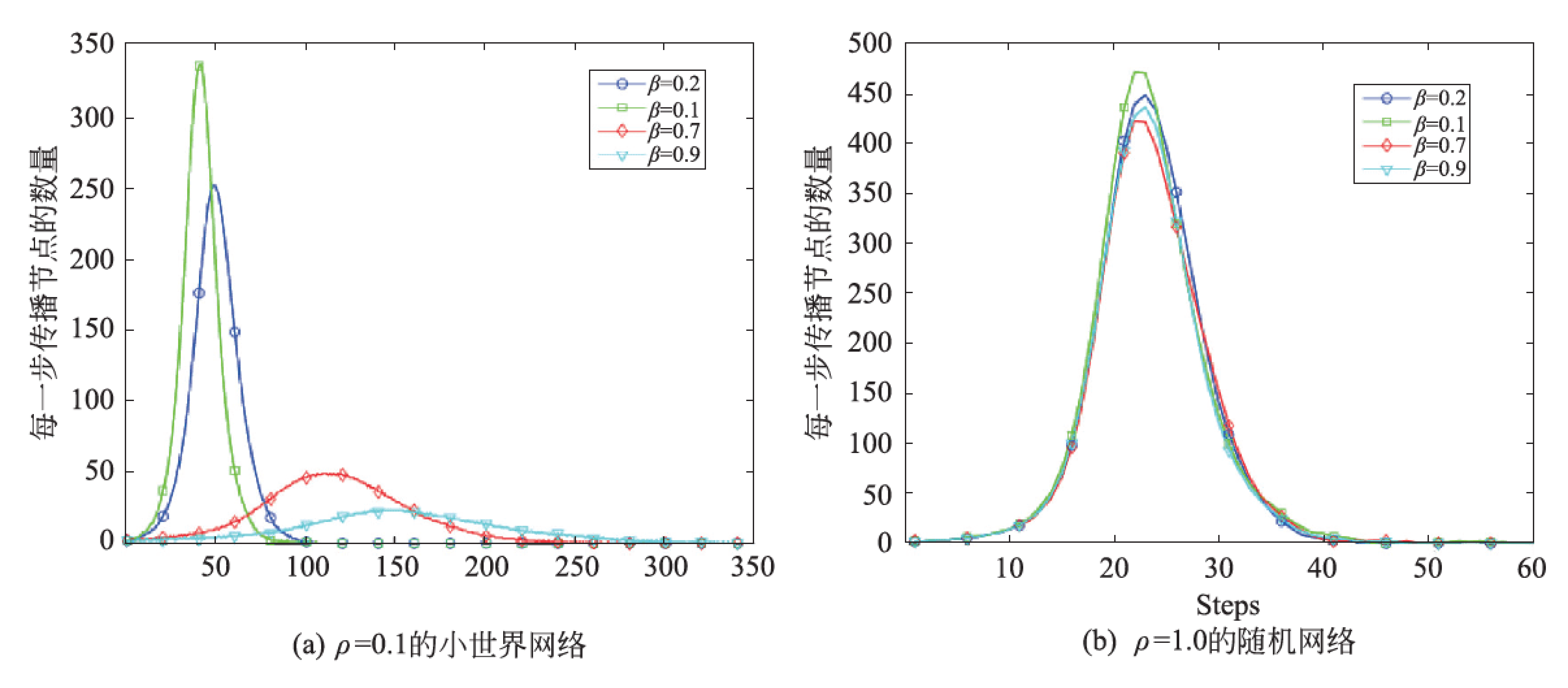
结果表明，当异质性很弱的时候，信息在小世界网络上的传播速度很慢。

然而，如图 5-3(a)所示，虽然异质性的加强降低了信息传播的速度，但却延长了该条信息的生命周期。在实际中，这可以解释为，当信息从一条不受信任的路径传递而来的时候，人们需要更多的时间去决定是否传播该条信息。

然而，并没有观测到异质性对随机网络的影响。如图 5-3(b)所示，对于给定的多个 β 值，传播的速度以及生命周期几乎都是一样的，这说明，异质性对随机网络几乎没有影响。

(4) 分析

最后，简单总结该一致性模型的所有实验结论。首先，异质性对于小世界网络的信息传播有着十分明显的影响。异质性越大，小世界的信息传播越被抑制。然而，对于随机网络(ρ 接近于 1 的时候)，异质性对它们的影响几乎没有。这种现象的背后主要有以下几种因素。首先，小世界网络中，边传播能力的不均匀性导致了信息容易被中断。假设对于一对邻居点 a 与 b，它们属于不同的社区。那么当异质性很强的时候，该条边的传播能力很差，信息很难通过这条边传播出去。然而，当异质性很弱的时候，这条边的传播能力与其他边相差无几，所以传播并没有受到抑制。其次，网络的拓扑结构很重要。f_{ij} 在模型中起着重要的作用，



(a) ρ=0.1的小世界网络　(b) ρ=1.0的随机网络

图 5-3　每一步传播节点的数量。其他参数是 $\alpha=8$，$K=6$，$N=10\,000$。实验结果取 1000 次独立实验的平均值

当异质性很强的时候，小世界网络对于该值十分敏感，因为 f_{ij} 的分布并不均匀；而对于随机网络，f_{ij} 的分布本身就比较均匀，所以结果对该值并不敏感。

7. **总结**

在该一致性模型中，研究了随机网络和小世界网络，哪种网络更适合信息传播。提出了一种考虑异质性的传播模型，并在此基础上进行了实验。结果表明，在不同的异质性强度下，模型所表现的结果完全不同。当异质性很强的时候，小世界网络的信息传播受到很大的抑制，传播的范围变窄，并被随机网络的传播范围超越。而且，随着异质性的加强，信息在小世界网络上传播的速度变慢，但其生命周期却延长了。当异质性很弱的时候，信息传播在小世界网络上的表现是最好的，并超越了随机网络与规则网络，这与文献[10,23]的结论一致。

然而，异质性对于随机网络并没有影响。在模型中，异质性体现在 q_{ij} 的计算上，而该值直接取决于 f_{ij}。在随机网络中，闭三角很罕见，任意两点之间的共同好友数都很少，且没有明显的区别性。这就决定了对于随机网络而言，f_{ij} 的值并没有区分性。所以，异质性对于随机网络并没用明显的影响。

然而，还有很多因素都可以影响信息传播。目前，对于个体上的研究还比较少。例如，当人们在决定转播一条信息的时候，是什么因素在该决策过程中起到了决策作用，是因为该信息的内容还是因为邻居的影响。这些都还是未知数，需要更多、更深入的研究。其次，只是通过一个参数来控制异质性的强弱，但并没有给一个衡量网络异质性强弱的方法。如何测量这种异质性的强弱，也是以后研究的方向。

5.2　竞争性信息传播

5.2.1　概述

随着社交网络的流行，人们的日常生活越来越离不开这些社交网络，随之产生的大量信

息也充斥着网络，无论是谣言还是商品广告信息等都会在网络上蔓延，因而了解信息背后的传播机理能够帮助人们更好地管理和控制网络上信息的传播。当一条信息在网络上传播扩散时，是否可以通过另一条竞争性的信息来抑制该信息的蔓延，或者通过协同的消息来促进该信息的传播？如果后发信息的目的是为了支持或抑制已有信息的传播，那么如何选择后发信息的时间点和空间节点以最大限度地帮助达到目的，将是本节讨论的重点。本节将研究两类信息同时传播的机理，以探索信息共同传播时的规律以及传播过程中各种因素的影响作用。通过对信息传播过程的模拟寻找竞争性信息传播过程的机理，从而帮助人们更好地对控制信息传播做出有效的决策。本节将从以下几个问题入手寻找竞争性信息传播的机理：

- 最终网络中传播各信息的节点分别是多少；
- 网络传播状态最终是否稳定；
- 强度大的信息会完全移除另一种信息，还是二者能共存。

对网络中的哪些特征能够影响两种信息的传播状态，我们将从以下几点加以分析：

- 传播网络的拓扑结构对信息传播是否有影响，如何影响；
- 网络中初始传播节点的度数对传播是否有影响，有何影响；
- 网络中初始传播节点的在线频率对传播是否有影响，有何影响；
- 网络中后发布的信息相对先发布的信息在发布时间上对最终网络传播状态有何影响。

5.2.2　竞争性信息传播模型

本节将介绍竞争性信息传播模型 ICSIS。模型中节点状态包括：未传播任何信息状态(susceptble)，表示为 S；传播 A 信息的状态，表示为 I_A；传播信息 B 的状态，表示为 I_B。初始状态为网络中节点的状态都为未传播任何信息的状态 S，继而信息 A 开始传播，信息 A 传播的起始是多源的，将多个节点作为传播源传播信息 A，在某一时刻，信息 B 也在网络中传播，两种类型的信息相互竞争和影响。两类信息同时在同一网络上传播，信息间存在依赖关系，即二者非独立，因此传播过程需考虑二者之间的相互作用关系。每个节点的状态转换是一个相对随机的过程，同时，下一次的状态与该节点的历史状态无关，只与当前状态有关，因此，整个传播过程可以看似一个马尔可夫随机过程。某个节点的状态动态变化过程可以表示成图 5-4 所示的过程。

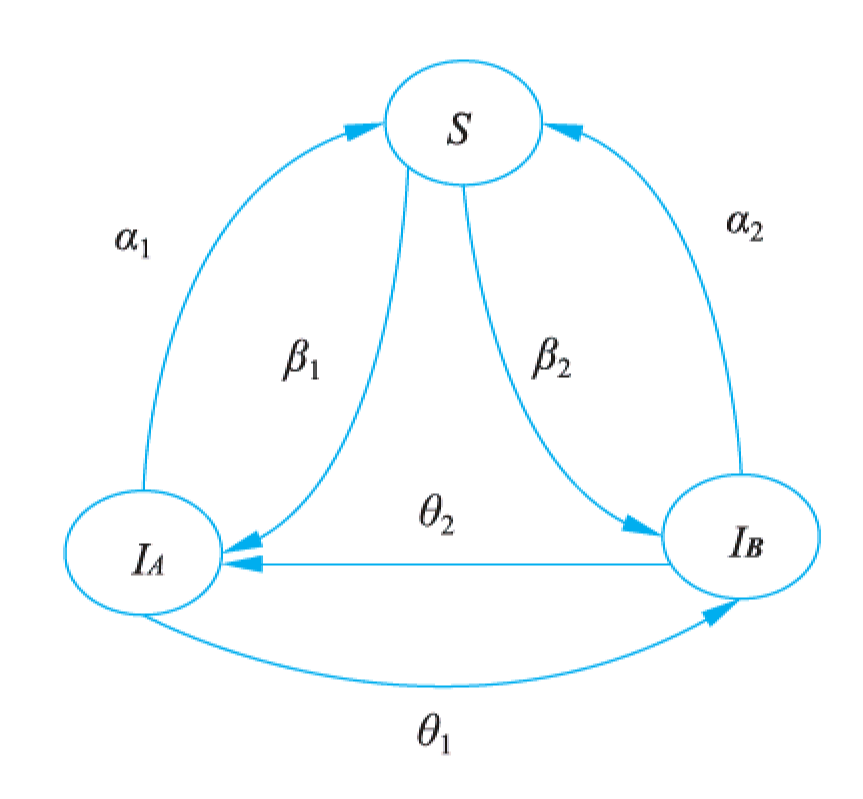


图 5-4　ICSIS 状态转移图

图 5-4 中，β_1 和 β_2 分别表示信息 A 和 B 的传播率，表示一个未传播状态下的节点对某类型信息的响应程度，传播率越高，表示该节点有更高的可能性去选择传播该条信息。α_1 和 α_2 分别表示为节点对信息 A 和 B 的遗忘率，随着时间的增加，节点会逐渐对传播过的信息遗忘，那么当前信息类型对该节点的影响力会随时间有所降低，若不考虑其他类型信息的影响，那么该节点最终又会变成未传播状态。θ_1 和 θ_2 分别表示信息类型 A 和信息类型 B 的相互影响力，θ_1 越大，则表示信息 B 的吸引力更大，能把传播信息类型 A 的节点状态改为传播信息类型 B，反之，同理。

5.2.2.1 模型相关变量解释

模型中用到了包括整个网络和节点内部的相关变量，这里，我们给出一个统一的定义。

对于整个网络有：

(1) $S(t)$：未传播任何消息的节点 t 时刻在网络中的总量；

(2) $I_A(t)$：传播 A 信息的节点 t 时刻在网络中的总量；

(3) $I_B(t)$：传播 B 信息的节点 t 时刻在网络中的总量；

(4) $\Theta(k)$：信息传播的网络的结构的度分布函数；

(5) $\dot{p}_{I_A,i}$：网络中第 i 个节点下一时刻是传播信息 A 的概率；

(6) $\dot{p}_{I_B,i}$：网络中第 i 个节点下一刻是传播信息 B 的概率；

(7) $\dot{p}_{s,i}$：网络中第 i 个节点下一时刻是未传播信息的概率；

(8) a_{ij}：网络 i，j 两个节点的连接情况，$a_{ij}=1$ 表示两节点间有连接，反之，$a_{ij}=0$ 表示两节点间没有连接。

对于传播过程，模型中有如下参数变量。

(1) β_1：信息 A 的传播率，β_2：信息 B 的传播率；

(2) α_1：信息 A 的遗忘率，α_2：信息 B 的遗忘率；

(3) θ_1：信息 A 的取代率，θ_2：信息 B 的取代率。

对于节点本身而言，有如下变量参数。

(1) $s_s^i(t)$：t 时刻 i 节点状态为未传播消息状态；

(2) $s_{I_A}^i(t)$：t 时刻 i 节点状态为传播信息 A；

(3) $s_{I_B}^i(t)$：t 时刻 i 节点状态为传播信息 B；

(4) $p_i^{I_A\to s}(t+1)$：下一时刻 i 节点从传播信息 A 完全恢复到未传播信息状态的概率；

$p_i^{I_A\to I_B}(t+1)$：下一时刻 i 节点从传播信息 A 状态被替代为传播信息 B 的概率；

$p_i^{I_A\to I_A}(t+1)$：下一时刻 i 节点保持传播信息 A 的状态的概率；

(5) $p_i^{I_B\to s}(t+1)$：下一时刻 i 节点从传播信息 B 完全恢复到未传播信息状态的概率；

$p_i^{I_B\to I_A}(t+1)$：下一时刻 i 节点从传播信息 B 状态被替代为传播信息 A 的概率；

$p_i^{I_B\to I_B}(t+1)$：下一时刻 i 节点保持传播信息 B 的状态的概率。

5.2.2.2 宏观 ICSIS 模型

所谓宏观即是从整个网络层面去构建整个 ICSIS 信息传播过程的模型，网络上节点的总量是稳定的，每个时刻变化的是网络中各个状态类型所占的比例。用 $S(t)$，$I_A(t)$，$I_B(t)$

分别表示 t 时刻，网络中节点各个状态对应的数量。用 $(s_s^i, s_{I_A}^i, s_{I_B}^i)^{\mathrm{T}}$ 表示一个节点的某时刻的状态，对于整个网络则有

$$\begin{cases} S(t) = \sum_{i=0}^{N} s_s^i(t) \\ I_A(t) = \sum_{i=0}^{N} s_{I_A}^i(t) \\ I_B(t) = \sum_{i=0}^{N} s_{I_B}^i(t) \end{cases} \tag{5-5}$$

根据平均场理论，ICSIS 信息传播模型可以表示成式(5-6)的微分方程组。

$$\begin{cases} \frac{\mathrm{d}S}{\mathrm{d}t} = -\beta_1 \times \frac{I_A(t)}{I_A(t)+S(t)} \times S(t) - \beta_2 \times \frac{I_B(t)}{I_B(t)+S(t)} \\ \frac{\mathrm{d}I_A}{\mathrm{d}t} = \beta_1 \times \frac{I_A(t)}{I_A(t)+S(t)} \times S(t) + \theta_1 \frac{I_A(t)}{I_A(t)+I_B(t)} \times I_B(t) - \theta_2 \frac{I_B(t)}{I_A(t)+I_B(t)} \times I_A(t) \\ \frac{\mathrm{d}I_B}{\mathrm{d}t} = \beta_2 \times \frac{I_B(t)}{I_A(t)+S(t)} \times S(t) + \theta_2 \frac{I_B(t)}{I_A(t)+I_B(t)} \times I_A(t) - \theta_1 \frac{I_A(t)}{I_A(t)+I_B(t)} \times I_B(t) \end{cases} \tag{5-6}$$

由于 $(s_s^i, s_{I_A}^i, s_{I_B}^i)^{\mathrm{T}}$ 是随机变量，因而有

$$\begin{cases} p_s(t) = E(s_s^i(t)) \\ p_{I_A}(t) = E(s_{I_A}^i(t)) \\ p_{I_B}(t) = E(s_{I_B}^i(t)) \end{cases} \tag{5-7}$$

则可以通过平均场理论得到概率的表示公式，如式(5-8)。

$$\begin{cases} \dot{p}_{I_A,i} = \beta_1 p_{s,i} \sum a_{ij} p_{I_A,j} + \theta_2 p_{I_B,i} \sum a_{ij} p_{I_A,j} - \theta_1 p_{I_A,i} \sum a_{ij} p_{I_B,j} - \alpha_1 p_{I_A,i} \\ \dot{p}_{I_B,i} = \beta_2 p_{s,i} \sum a_{ij} p_{I_B,j} + \theta_1 p_{I_A,i} \sum a_{ij} p_{I_B,j} - \theta_2 p_{I_B,i} \sum a_{ij} p_{I_A,j} - \alpha_2 p_{I_B,i} \\ \dot{p}_{s,i} = \alpha_1 p_{I_A,i} + \alpha_2 p_{I_B,i} - \beta_1 p_{s,i} \sum a_{ij} p_{I_A,j} - \beta_2 p_{s,i} \sum a_{ij} p_{I_B,j} \end{cases} \tag{5-8}$$

由归一化条件有

$$\dot{p}_{s,i} + \dot{p}_{I_A,i} + \dot{p}_{I_B,i} = 1 \tag{5-9}$$

则式(5-8)可以简化为

$$\begin{cases} \dot{p}_{I_A,i} = \beta_1 (1 - p_{I_A,i} - p_{I_B,i}) \sum a_{ij} p_{I_A,j} + \theta_2 p_{I_B,i} \sum a_{ij} p_{I_A,j} - \theta_1 p_{I_A,i} \sum a_{ij} p_{I_B,j} - \alpha_1 p_{I_A,i} \\ \dot{p}_{I_B,i} = \beta_2 (1 - p_{I_A,i} - p_{I_B,i}) \sum a_{ij} p_{I_B,j} + \theta_1 p_{I_A,i} \sum a_{ij} p_{I_B,j} - \theta_2 p_{I_B,i} \sum a_{ij} p_{I_A,j} - \alpha_2 p_{I_B,i} \end{cases} \tag{5-10}$$

5.2.2.3　微观 ICSIS 模型

5.2.2.2 节是从宏观角度得出一个网络多条信息传播的模型和概率模型，而若从微观角度，即从某一个节点每一时刻的状态转化的规律进行研究的话，需考虑影响某个节点状态变化的因素和变化的规则。如宏观状态下描述的一样，一个节点状态转换的影响因素包括：某信息类型的传播率（可以认为是节点对信息的响应度）、信息的遗忘率（该节点对该信息的

热情度和记忆的衰减率）以及其他类型信息的干扰，而对于节点自身特征而言，其在线频率与节点的度数则是影响传播的内因（在线频率表现为该节点的活跃度，在社交网络中体现为用户在线和分享的活跃程度，越活跃则越有可能传播信息，节点度数则体现为用户在社交网络中好友的数量，度数多的可以认为是好友比较多）。节点度数的属性为空间上的属性，在讨论时间上的传播规律时不考虑这个因素，在讨论空间对传播规律的影响时将会提到。那个节点的在线频率用一个概率来表示：$p_f, p_f \in [0,1]$。对于网络节点的状态转换满足一定的规则，即两条竞争性信息在网络上传播符合一定的规则，这些规则与影响因子有关。

当节点的邻居节点有信息 A 传播状态的时候，那么该节点有一定的概率传播信息 A，表示为式(5-11)。

$$p_{I_A}(t+1) = v_i . p_f\left(1-\beta_1\prod_{j}^{N} a_{ij}v_j^{I_A}(t)\right) \tag{5-11}$$

同理，当节点的邻居节点有信息 B 传播状态时，传播信息 B 的概率为

$$p_{I_B}(t+1) = v_i . p_f\left(1-\beta_2\prod_{j}^{N} a_{ij}v_j^{I_B}(t)\right) \tag{5-12}$$

若一个节点同时满足邻居节点有信息 A 和信息 B 的传播状态，那么都存在着一定概率传播其中的任一类型信息，这种情况下考虑任一类型信息对该节点的吸引力，加上对该节点的单独传播概率，综合考虑如下。

$$\text{compare}(\lambda_1 p_{I_A}(t), \lambda_2 p_{I_B}(t))$$

那么影响力更大的信息类型将成为该节点的下一时刻状态。

若节点状态为已传播信息的状态，那么满足以下规则。节点传播状态为传播信息 A，那么下一个状态有可能回到不传播状态，当然，这个过程应该是一个需要时间的过程（现实中存在这种现象，就是过段时间有些人会分享些曾经分享过的东西，原因是因为忘记曾经分享过）。根据艾宾浩斯遗忘规律，节点遗忘对该类型信息传播满足一个曲线规律，但也有可能一个点选择重复传播（现实中存在这种现象，人们会反复分享同类信息，如广告、微商等，目的是多次传播引起他人关注），那么该节点就存在着这样一个概率，即让该节点重新变回未传播任何状态的节点，以便于能够再次传播该类型信息，同时又考虑到一个节点虽然当前传播状态为 A 类信息，但可能接到信息 B 时，对 B 的感兴趣程度要高于前一信息类型，那么该节点便会很快变成传播信息 B 的状态，综合考虑，表示为

$$\begin{cases} p_i^{I_A\to s}(t+1) = v_i . p_f\alpha_1 p_{I_A,i}(t) + p_{s,i}(t) \\ p_i^{I_A\to I_B}(t+1) = \dfrac{\lambda_2}{\lambda_1}v_i . p_f\left(1-\prod(1-\theta_1 a_{ij}s_{I_B}^j)\right)p_{I_A,i}(t) \\ p_i^{I_A\to I_A}(t+1) = p_{I_A,i} - v_i . p_f\alpha_1 p_{I_A,i}(t) \end{cases} \tag{5-13}$$

Compare$(p_i^{I_A\to s}, p_i^{I_A\to I_B}, p_i^{I_A\to I_A})$，概率最大者的点将成为该状态。

同理，对于当前传播信息 B 类型的节点，有

$$\begin{cases} p_i^{I_B\to s}(t+1) = v_i . p_f\alpha_2 p_{I_B,i}(t) + p_{s,i}(t) \\ p_i^{I_B\to I_A}(t+1) = \dfrac{\lambda_1}{\lambda_2}v_i . p_f\left(1-\prod(1-\theta_2 a_{ij}s_{I_A}^j)\right)p_{I_B,i}(t) \\ p_i^{I_B\to I_B}(t+1) = p_{I_B,i} - v_i . p_f\alpha_2 p_{I_B,i}(t) \end{cases} \tag{5-14}$$

Compare($p_i^{I_B\to s}$, $p_i^{I_B\to I_A}$, $p_i^{I_B\to I_B}$)，概率最大者为该节点下一时刻状态。

这样，在微观模型中，公式如式(5-15)所示。

$$\begin{cases}\dot{p}_{I_A,i}=\left[\left(1-\prod(1-\beta_1 p_f a_{ij} s_{I_A}^j)\right)-\lambda_2\left(1-\prod(1-\beta_2 p_f a_{ij} s_{I_B}^j)\right)\right]p_{s,i}-\alpha_1 p_{I_A,i}\\ \qquad -\left(1-\prod(1-\theta_1 p_f a_{ij} s_{I_A}^j)\right)p_{I_A,i}+\left(1-\prod(1-\theta_2 p_f a_{ij} s_{I_B}^j)\right)p_{I_B,i}\\ \dot{p}_{I_B,i}=\left[\left(1-\prod(1-\beta_2 p_f a_{ij} s_{I_B}^j)\right)-\lambda_1\left(1-\prod(1-\beta_1 p_f a_{ij} s_{I_A}^j)\right)\right]p_{s,i}-\alpha_2 p_{I_B,i}\\ \qquad -\left(1-\prod(1-\theta_2 p_f a_{ij} s_{I_B}^j)\right)p_{I_B,i}+\left(1-\prod(1-\theta_1 p_f a_{ij} s_{I_A}^j)\right)p_{I_A,i}\\ \dot{p}_{s,i}+\dot{p}_{I_A,i}+\dot{p}_{I_B,i}=1\end{cases} \tag{5-15}$$

公式中其他字母的含义如上面所述，λ_1 和 λ_2 表示同一时刻一个节点能同时接收到两个类型时对某类型信息的偏好概率，有

$$\lambda_1+\lambda_2=1 \tag{5-16}$$

其中，未传播任何消息节点的概率值可以由传播两个类型信息的概率导出，因此，公式(5-16)可以简化为只有两个变量的形式，如式(5-17)所示。

$$\begin{cases}\dot{p}_{I_A,i}=\left[\left(1-\prod(1-\beta_1 p_f a_{ij} s_{I_A}^j)\right)-\lambda_2\left(1-\prod(1-\beta_2 p_f a_{ij} s_{I_B}^j)\right)\right](1-p_{I_A,i}-p_{I_B,i})\\ \qquad -\alpha_1 p_{I_A,i}-\left(1-\prod(1-\beta_1 p_f a_{ij} s_{I_A}^j)\right)p_{I_A,i}+\left(1-\prod(1-\theta_2 p_f a_{ij} s_{I_B}^j)\right)p_{I_B,i}\\ \dot{p}_{I_B,i}=\left[\left(1-\prod(1-\beta_2 p_f a_{ij} s_{I_B}^j)\right)-\lambda_1\left(1-\prod(1-\beta_1 p_f a_{ij} s_{I_A}^j)\right)\right](1-p_{I_A,i}-p_{I_B,i})\\ \qquad -\alpha_2 p_{I_B,i}-\left(1-\prod(1-\theta_2 p_f a_{ij} s_{I_B}^j)\right)p_{I_B,i}+\left(1-\prod(1-\beta_1 p_f a_{ij} s_{I_A}^j)\right)p_{I_A,i}\end{cases} \tag{5-17}$$

5.2.3 竞争性信息传播模型稳定性分析

本节通过解得的平衡点对 ICSIS 信息传播模型进行稳定态分析。首先介绍稳定性分析过程中用到的雅可比矩阵和 Routh-Hurwitz 定理，然后推出满足 ICSIS 信息传播模型稳定的条件。

所谓的稳定性，是指系统受到的扰动作用消除后，一段过渡过程以后能否回复到原来的平衡状态或足够准确地返回到之前的平衡态。如果系统能够恢复到此前的平衡状态，那么称该系统是稳定的；若扰动消失后系统不能恢复到原来的平衡状态，反而偏差变得更大，则称该系统是不稳定的。系统的稳定性一般分为两种情况：一种是小范围稳定，起始的偏差需要在一定的限度内系统才能稳定，超出了这个限定范围则不再稳定；而另一种是大范围稳定，这种情况起始的偏差可以很大，但系统仍然稳定。本节所讨论的 ICSIS 模型的稳定性是一个全局的稳定性问题。

5.2.3.1 Routh-Hurwitz 定理

在控制系统理论中，Routh-Hurwitz 稳定性判据是线性非时变控制系统稳定的充分必要条件。Routh 测试是英国数学家爱德华·约翰·劳斯在 1876 年提出的一种有效的递归算法，用以判断线性系统的特征多项式的所有根中是否有负根。德国数学家阿道夫·赫维兹在 1895 年提出将多项式的系数放在一个方阵中(这个方阵后来命名为赫维兹矩阵)，多项

式是稳定的当且仅当赫维兹矩阵的子阵的行列式都是正的。两种方法是等效的。Routh 测试提供了一个更加有效的方式计算赫维兹行列式而非直接计算。若一个多项式满足 Routh-Hurwitz 标准，则称该多项式为赫维兹多项式。线性系统是否稳定取决于该系统微分方程的特征多项式的根，若特征多项式的全部根都是负实数或实部为负的复数，则该系统是稳定的。

5.2.3.2 稳定性分析

网络达到平衡点，那么网络将处于平衡态。因此有

$$\begin{cases}\dfrac{\mathrm{d}S(t)}{\mathrm{d}t}=0\\[2mm]\dfrac{\mathrm{d}I_A(t)}{\mathrm{d}t}=0\\[2mm]\dfrac{\mathrm{d}I_B(t)}{\mathrm{d}t}=0\end{cases}\tag{5-18}$$

结合式(5-6)和式(5-18)，得到

$$\begin{cases}-\beta_1\Theta(k)S(t)I_A(t)-\beta_2\Theta(k)S(t)I_B(t)+\alpha_1I_A(t)+\alpha_2I_B(t)=0\\-\alpha_1I_A(t)-\theta_1\Theta(k)I_A(t)I_B(t)+\beta_1\Theta(k)S(t)I_A(t)+\theta_2\Theta(k)I_B(t)I_A(t)=0\\-\alpha_2I_B(t)-\theta_2\Theta(k)I_B(t)I_A(t)+\beta_2\Theta(k)S(t)I_B(t)+\theta_1\Theta(k)I_A(t)I_B(t)=0\end{cases}\tag{5-19}$$

设平衡点 $B=(S,I_A,I_B)^{\mathrm{T}}$，求解上式，可以得到 $B_0=(1,0,0)^{\mathrm{T}}$ 是上面方程组的一个解，则 $B_0=(1,0,0)^{\mathrm{T}}$ 是 ICSIS 信息传播模型的一个平衡点，达到平衡点后，网络信息传播属于一个动态平衡。

定理 5.1 当 $R_0<1$ 时，平衡点 $B_0=(1,0,0)^{\mathrm{T}}$ 是全局渐进稳定的。

证明：

式(5-19)的雅可比矩阵为

$$J=\begin{bmatrix}-\beta_1\Theta(k)I_A-\beta_2\Theta(k)I_B & -\beta_1\Theta(k)S+\alpha_1 & -\beta_2\Theta(k)S+\alpha_2\\ \beta_1\Theta(k)I_A & -\alpha_1-\theta_1\Theta(k)I_B+\beta_1\Theta(k)S & \theta_2\Theta(k)I_A\\ \beta_2\Theta(k)I_B & \theta_1\Theta(k)I_B & -\alpha_2-\theta_2\Theta(k)I_A+\beta_2\Theta(k)S\end{bmatrix}\tag{5-20}$$

在平衡点 $B_0=(1,0,0)^{\mathrm{T}}$ 处，$J(B_0)$ 为

$$J(B_0)=\begin{bmatrix}0 & -\beta_1\Theta(k)+\alpha_1 & -\beta_2\Theta(k)+\alpha_2\\0 & -\alpha_1+\beta_1\Theta(k) & \theta_2\Theta(k)\\0 & \theta_1\Theta(k) & -\alpha_2+\beta_2\Theta(k)\end{bmatrix}\tag{5-21}$$

设矩阵的特征多项式为

$$\lambda^3+k_2\lambda^2+k_1\lambda+k_0=0\tag{5-22}$$

式(5-21)的多项式求解如下：

$$|J(\lambda I-B_0)|=\begin{vmatrix}\lambda & -\beta_1\Theta(k)+\alpha_1 & -\beta_2\Theta(k)+\alpha_2\\0 & \lambda-\alpha_1+\beta_1\Theta(k) & 0\\0 & 0 & \lambda-\alpha_2+\beta_2\Theta(k)\end{vmatrix}=0\tag{5-23}$$

则有

$$\lambda(\lambda-\alpha_1+\beta_1\Theta(k))(\lambda-\alpha_2+\beta_2\Theta(k))=0 \tag{5-24}$$

解得

$$\begin{cases} k_2=\alpha_1+\alpha_2-\beta_1\Theta(k)-\beta_2\Theta(k) \\ k_1=(\alpha_1-\beta_1\Theta(k))(\alpha_2-\beta_2\Theta(k)) \\ k_0=0 \end{cases} \tag{5-25}$$

因为

$$R_0=\frac{\beta_1\Theta(k)}{\alpha_1}<1,\quad 或\ R_0=\frac{\beta_2\Theta(k)}{\alpha_2}<1$$

所以

$$\begin{cases} k_2>0 \\ k_1>0 \\ k_0=0 \end{cases} \tag{5-26}$$

根据 Routh-Hurwitz 稳定性判据[19]，B_0是稳定点。 □

当$\frac{\beta_1}{\alpha_1}<\frac{1}{\Theta(k)}$，$\frac{\beta_2}{\alpha_2}<\frac{1}{\Theta(k)}$时，系统处于问题态，信息 A 与信息 B 的传播都为 0，为了能够在网络上有信息传播，本节模拟条件设定为$\frac{\beta_1}{\alpha_1}>\frac{1}{\Theta(k)}$，$\frac{\beta_2}{\alpha_2}>\frac{1}{\Theta(k)}$，那么信息 A 与信息 B 将会在网络上传播，最终达到一个非零平衡点，使网络重新达到稳定态，具体的稳定态在模拟实验结果中给出。

5.2.4　竞争性信息传播模型模拟

在 ICSIS 模型中，网络的拓扑结构以及节点在线频率等因素都是整个网络状态分布的重要的影响因子，随着这些影响因子的变化，每个迭代时刻的网络节点状态分布将会随之发生改变，通过对不同影响因子的控制，模拟 ICSIS 模型的传播过程，观察传播过程中具体的网络状态分布。因为信息的传播过程可以看作是一个马尔可夫过程，所以在模拟实验中，带有许多随机过程，模拟结果需要取通过多次实验结果加起来的平均，这样即能有效地避免意外的随机结果对整个实验结果产生的影响。

本节介绍模拟 ICSIS 模型信息传播过程的算法，由第 3 章的模型推导可以知道，对于网络中的每个节点，其下一步的状态受到的影响因素有哪些以及构成的函数，其次可以得知两种信息是如何在一个节点上形成竞争以及竞争节点上下一时刻的状态的决定机制。ICSIS 模型提供了这些信息传播过程的一些规则，将这些规则应用到传播过程的模拟实验中，输出传播过程中每个时刻网络状态的情况。

在整个传播过程中，节点的状态发生着动态变化，每一时刻节点拥有某一状态，那么不同的状态在往下一状态变化时其处理过程是不同的，下面将分别介绍不同状态的节点的变化过程算法。

(1) 当节点状态为未传播信息状态时，其下一刻的状态变化规则如表 5-2 所示。

表 5-2　未传播信息状态节点的变化规则

Rule 5.1　doWithCaseSus(G graph)

```
Input: G graph
Output: stateEventually
for j←i+1 to maxnodenum do
    if adjMatrix(i,j)=true then   //adjMatrix represents adjacent metrix
        if vj.state=infectByOne then
            infectedByOne++   //count the adjacent nodes with preading msgA
        else if vj.state=infectByTwo then
            infectedByTwo++   //count the adjacent nodes with preading msgB
if infectedByOne>0 and infectedByTwo=0 then   //only receive info. Type A
    onlyTypeOne()
else if infectedByOne=0 and infectedByTwo>0 then   //only receive inf. Type B
    onlyTypeTwo()
else if infectedByOne>0 and infectedByTwo>0 then //receive two type
    twoTypes()
```

节点有可能在某一时刻接收不到任何消息，那么状态将保持；也有可能只接收到某一类型的信息，然后以一定的概率传播，还有可能同时接收到两类信息，那么将对二者的传播可能性进行讨论和比较，具体如下处理。

① 当只接收到信息 *A* 时，处理规则如表 5-3 所示。

表 5-3　只接收信息 *A* 的处理规则

Rule 5.1.1　onlyTypeOne(G graph)

```
Input: G graph
Output: refreshed graph
p_IA←v_i.frequant * (1-(1-β)^infectedbyone)
p_rand←random(0,1)
if p_rand<p_IA then
    v_i.state=infecteByOne
    I_A.add(S.delect(vi))
    v_i.infectlevel1=p_IA
```

② 当只接收到信息 *B* 时，处理规则如表 5-4 所示。

表 5-4　只接收信息 *B* 的处理规则

Rule 5.1.2　onlyTypeTwo(G graph)

```
Input: G graph
Output: refreshed graph
p_IB←v_i.frequant * (1-(1-β)^infectedbytwo)
p_rand←random(0,1)
if p_rand<p_IB then
    v_i.state=infecteByTwo
    I_B.add(S.delect(vi))
    v_i.infectlevel1=p_IB
```

③ 当同时接收到两类信息时，需要进行更细致的处理，对两类型信息的传播可能性进行比较，同时考虑信息的吸引力，具体规则如表 5-5 所示。

表 5-5　两类型同时接收时转换规则

Rule 5.1.3　twoTypes(G graph)

```
Input: G graph
Output: refreshed graph
p_IA ← v_i.frequant * (1-(1-β)^infectedByOne)
p_IB ← v_i.frequant * (1-(1-β)^infectedByTwo)
p_rand ← random(0,1)
if p_IA > p_rand and p_IB < p_rand then
    v_i.state ← infecteByOne
    I_A.add(S.delect(vi))
    v_i.infectlevel1 ← p_IA
else if p_IA < p_rand and p_IB > p_rand then
    v_i.state ← infecteByTwo
    I_B.add(S.delect(vi))
    v_i.infectlevel2 ← p_IB
else if p_IA > p_rand and p_IB > p_rand then
    if p_IA * λ1 > p_IB * λ2 then
        v_i.state ← infecteByOne
        I_A.add(S.delect(vi))
        v_i.infectlevel1 ← p_IA
    else
        v_i.state ← infecteByTwo
        I_B.add(S.delect(vi))
        v_i.infectlevel2 ← p_IB
```

(2) 当节点当前状态为传播 *A* 信息时，该节点有一定的遗忘率使其恢复到传播的状态，同时还有一定的概率给新的信息取代覆盖，具体转换规则如表 5-6 所示。

表 5-6　传播信息 *A* 节点状态处理规则

Rule 5.2　doWithCaseIa(G graph)

```
Input: G graph
Output: newState
for j ← i+1 to maxnodenum do    //find msgB adjacent nodes
    if adjMatrix(i,j) = true and v_j.state = infectByTwo then
        infecteTwoNum++
p_s ← v_i.frequant * α1 * v_i.infectlevel1
p_IB ← (λ2/λ1) * (1-(1-θ1)^infectedByTwo) * v_i.frequant
```

(3) 若当前状态为传播信息 *B*，那么其转换规则与第(2)点中的类似，在具体的转化上其区别在于考虑不同的邻接点状态，具体的转化规则如表 5-7 所示。

表 5-7　传播信息 *B* 节点状态处理规则

Rule 5.3　doWithCaseIb(G graph)

Input：G graph
Output：newState
1. **for** j←i+1 **to** maxnodenum **do**　//find msgA adjacent nodes
2. 　**if** adjMatrix(i,j)=true **and** vj. state=infectByOne **then**
3. 　　infecteOneNum++
4. $p_s \leftarrow v_i.frequant * \alpha_2 * v_i.infectlevel2$
5. $p_{IA} \leftarrow \frac{\lambda_1}{\lambda_2} * (1-(1-\theta_2)^{infectedByOne}) * v_i.frequant$
6. $p_{IB} \leftarrow v_i.infectlevel2 - p_s$
7. $p \leftarrow compare(p_s, p_{IA}, p_{IB})$
8. newState←Changestate(i,p)

从表 5-2～表 5-7，详细描述了节点状态的一些具体规则，通过这些规则，对整个传播模型进行传播过程的模拟实验，具体的传播过程模拟算法如表 5-8 所示。

表 5-8　传播过程模拟算法

算法 5.1　ICSIS_simulation()

Input：G=(V,E)
Output：spreading process
Begin：
1. $S \leftarrow \varnothing, I_A \leftarrow \varnothing, I_B \leftarrow \varnothing$，　//init the state sets
2. **for each** $v_i \in V$　//init the graph to spreading
　a) v_i. state←susceptable，S. add(vi)
　b) v_i. frequant←random(0,1)
3. $I_A \leftarrow$(selectNodeFrom(V,indexs)),S. delete(v_{indexs})　//select some nodes from V，value them with state spread type A information
4. $I_B \leftarrow$(selectNodeFrom(V,indexs)),S. delete(v_{indexs})　//the same as 3
5. loop←0
6. **While** true **do**　//start iteration
　loop++
7. **for** i←0 **to** maxnodenum **do**　//maxnodenum denotes numbers of nodes in G. V
8. 　Switch v_i. state
9. 　　Case susceptable：
10. 　　　**doWithCaseSus**()
11. 　　Case infectByone：
12. 　　　**doWithCaseIa**()
13. 　　Case infectByTwo
14. 　　　**doWithCaseIb**()
15. **if** I_A. len≫I_B. len **or** I_A. len≪I_B. len **then**
16. 　return(loop，S,I_A,I_B)
End spreading process

5.2.5　网络拓扑结构对竞争性信息传播的影响

在 ICSIS 信息传播模型中，在状态模型和概率模型方面都涉及网络的拓扑结构特征，$\Theta(k)$是网络节点的度分布函数，复杂网络中的随机网络度分布满足泊松分布，而无标度的复杂网络模型的度分布则呈幂律分布。由于网络拓扑存在的差异，网络图结构上表现为邻接矩阵的特性存在着差异，如此一来，节点状态的分布与最终的稳定态都会有所差异。本节通过对不同的复杂网络拓扑结构分别用程序模拟传播过程，然后找出每种网络拓扑对应的传播特征，相互比较，从而研究网络拓扑结构对传播过程的影响。

5.2.5.1　小世界网络传播模拟

小世界网络(small-world networks)模型用来生成现实世界的人际网络或社交网络的一些特征。一方面，小世界网络每个节点间的平均距离非常短，另一方面，小世界网络具有高聚类系数。同时，小世界网络还是一个可以扩展的熵的系数网络。随着少数随机链路加入到结构化的网络上，平均路径长度会很快减少。节点间的平均路径长度呈 $\log N$ 的序列分布，那么该网络体现了小世界的特征。然而，WS 小世界网络(Watts-Strogatz small world networks)同时表现出了小世界和高聚类的属性。本节模拟的网络结构拟用 WS 小世界网络，该网络的构建算法如表 5-9 所示。

表 5-9　小世界网络构建算法

算法 5.2　WSNetwork(Graph，reconEdges，recon_p)

```
输入：初始化图 G0：graph
      选择重连的边数：reconEdges
      选择重连的概率：recon_p
输出：小世界网络图 wsGraph
1. wsGraph = graph  //init the wsgraph
2. For each cur_node in wsGraph do
3.   For i:=0 to reconEdges do
4.     Link(cur_node, node(i))  //link every k neiberough nodes
5. For each cur_node in wsGraph do
6.   For i:=0 to reconEdges do
7.     Randp:= getRandom()
8.     If randp<recon_p then
9.       deLink(cur_node,node(i))
10.      newNode:=getNodeFromUnLinked(unlink(cur_node))
11.      Link(cur_node, newNode)
```

本节的模拟实验中，网络的总节点数设为 $N=4000$，每次重连的边数 reconEdges=100，每次重连的概率为 recon_p=0.6，构建的 WS 小世界网络中的节点度数形成以重连边数为中心向两边递减的分布。很少具有高度数的节点，则该网络的节点度分布如图 5-5 所示。

小世界网络的度分布呈现以重连边数为中心的钟形正态分布。在上面构建的 WS 小世

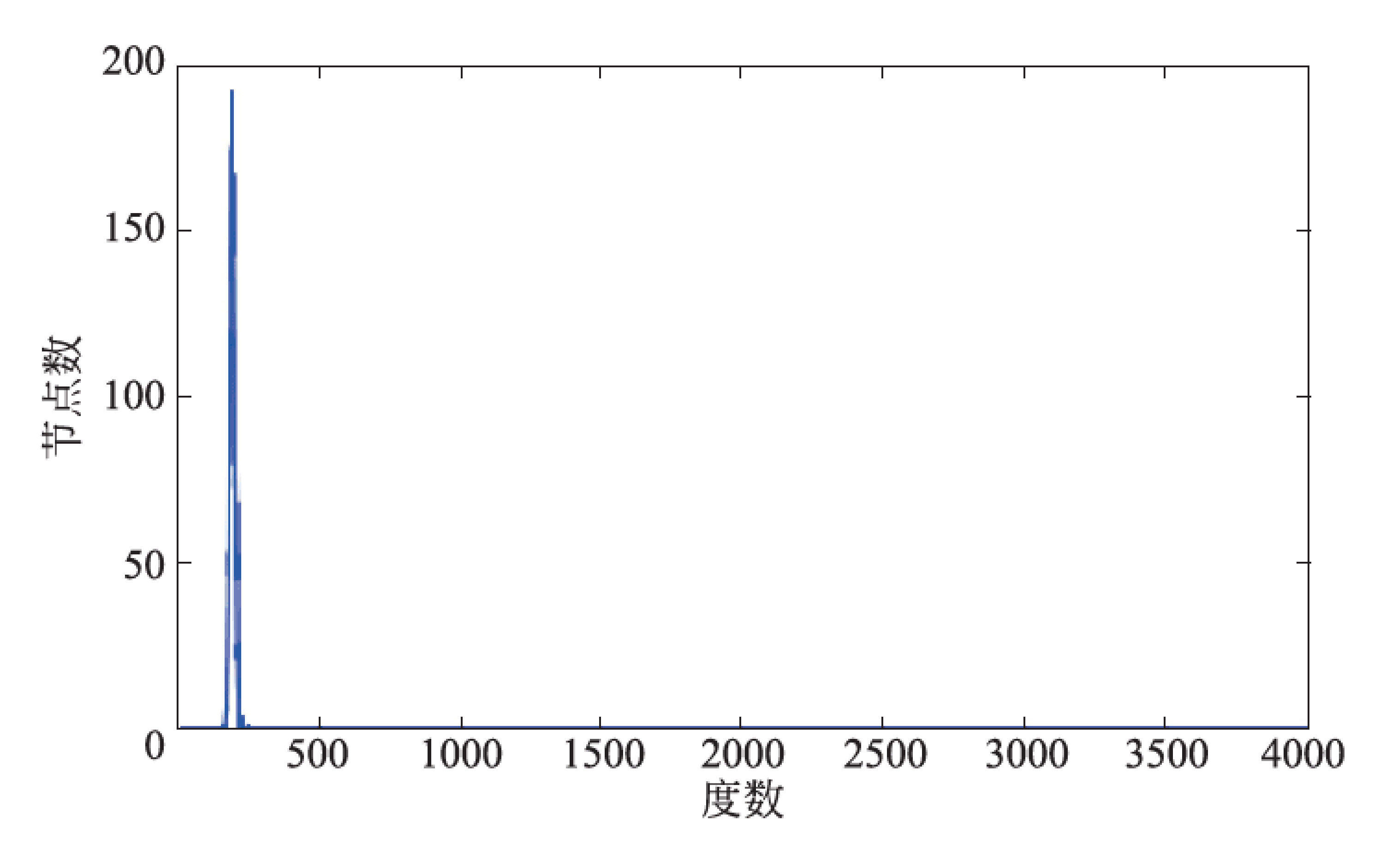


图 5-5　WS 小世界网络度分布

界网络上进行模拟 ICSIS 信息传播模型，在模拟实验中，信息类型 B 是后发者，在信息 A 发布 5 个迭代的单位时间后随机从网络中多源传播。模拟实验的条件设定如表 5-10 所示。

表 5-10　小世界网络上信息传播过程模拟条件设定

参　数	设定值
信息 A 传播率 β_1	0.5
信息 B 传播率 β_2	0.6
信息 A 遗忘率 α_1	0.1
信息 B 遗忘率 α_2	0.1
信息 A 对 B 的取代率 θ_1	0.4
信息 B 对 A 的取代率 θ_2	0.6
信息 A 的吸引力 λ_1	0.4
信息 B 的吸引力 λ_2	0.6
信息 A 的发布时间 t_1	0
信息 B 的发布时间 t_2	5

在模拟实验的条件设定中，信息 B 的强度大于信息 A，也就是说，在传播过程中，信息 B 更具侵略性，有更大的可能性覆盖信息 A 的传播，这样设定的目的是为了观察强度更大的信息 B 是否会覆盖强度稍弱的信息 A，若是，是如何覆盖的，网络拓扑结构在这之中起到什么作用。按照表 5-10 所示的传播过程参数设定，模拟信息在 *WS* 小世界网络上的传播过程，当网络信息传播达到动态平衡时，即节点连续迭代变化不大时，停止传播，输出结果，则传播过程中每个迭代单位网络所有节点状态的变化如图 5-6 所示。

由于信息 B 的强度大于信息 A 的强度，因此在很大程度上信息 B 能够覆盖信息 A，随着迭代时间的延长，信息 B 的传播逐渐占据整个网络。信息 B 开始传播前只有信息 A 在网络中传播，当信息 A 快要覆盖大部分网络时，信息 B 开始多源传播，很快便抑制住了信息 A 的传播。同时，自身传播覆盖大部分网络。最终网络达到了一个稳定的状态，说明传播开始后存在一个全局平衡点使网络达到稳定态。

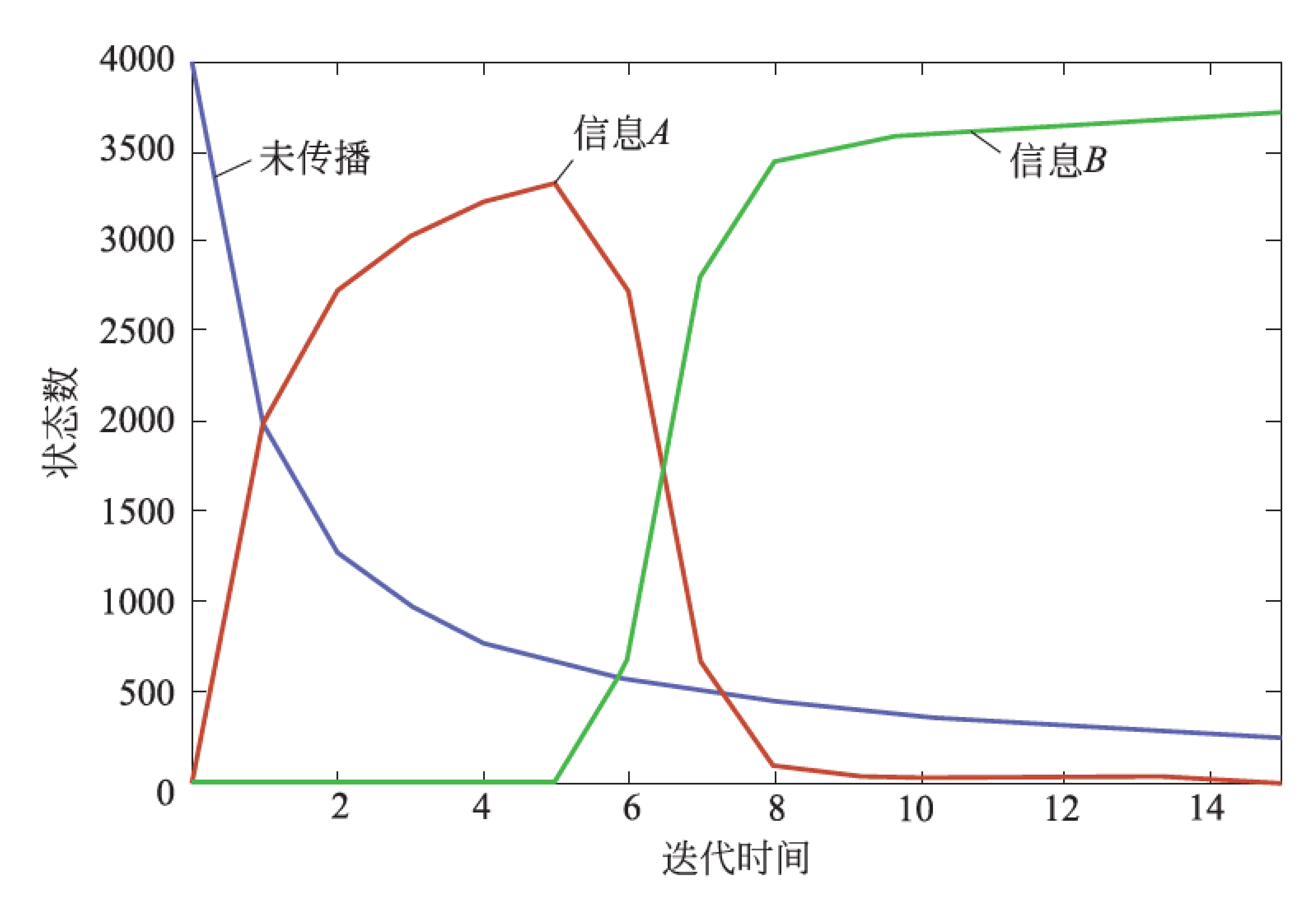


图 5-6　信息传播过程网络状态变化

5.2.5.2　BA 无标度网络传播模拟

网络度分布呈幂律分布的网络称为无标度网络(scale-free network),即有

$$P_k \propto k^{-\gamma}$$

其中,k 表示网络中有 k 个连接的节点,即度数为 k 的节点的集合;P_k 表示节点度数为 k 的概率;无标度网络中,通常 γ 满足($2<\gamma\leqslant 3$)。无标度网络的典型特点是网络中的大部分节点只有与很少的其他节点连接,而只有小部分节点与非常多的其他节点连接。这些能与非常多的网络中其他节点连接的节点称为集散节点,它们的存在使得无标度网络对意外故障有强大的承受能力,但在面对协同性进攻时,则相对比较脆弱。Albert-László Barabási 和 Réka Albert 在 1999 年的一篇论文中提出了一个用来解释复杂网络无标度特性的模型,叫作 BA 模型(Barabási-Albert model),BA 模型可以用来模拟许多现实生活中的网络属性。该模型基于如下两个假设:

(1) 许多现实网络是不断扩大和不断增长而来的;

(2) 新的节点在加入网络时更倾向于连接到更多连接的节点上。

本节的实验模拟用 BA 算法构建一个无标度网络,BA 模型的构建算法如表 5-11 所示,网络从开始的 m_0 个节点的完全连接图拓展,每次加入 $m\leqslant m_0$ 个新的连接。

表 5-11　BA 网络构建算法

算法 5.3　BANetwork(Graph, startNum, newLinks).
输入:初始化图 G_0:graph 起始网络节点数量:startNum 每次加入网络的新连接数:newLinks **输出**:BA 网络图 baGraph 1. baGraph=start_graph　//init the bagraph 2. pvec={},nodevex={} 3. **For** each cur_node **in** start_node **do** 4. **For** i:=0 **to** startNum **do**

续表

```
            Link(cur_node, node(i))   //construct complete graph
            nodevec.add(cur_node)
Do
      newNode:=createNode()
      For idx:=0 to nodevec.len do
            pvec[idx]=nodevec[idx].degree/totaldegree
       For j:=0 to newLinks do
            Randp:=getRandomn()
            For k:=0 to nodevec.len do
                  p:=p+pvec[j]
                  if randp<p then
                        Link(newNode, nodevec[j])
            nodevec.add(newNode)
            totaldegree:=totaldegree+2 * newLinks
When nodevec.len<N
```

在本节的模拟实验中，BA 网络的构建中起始的节点数设为 $m_0=50$，每次新加的连接数量 $m=40$，网络总节点数设为 $N=4000$，则该 BA 无标度网络的节点度分布呈一个幂律分布，具体的度分布如图 5-7 所示。

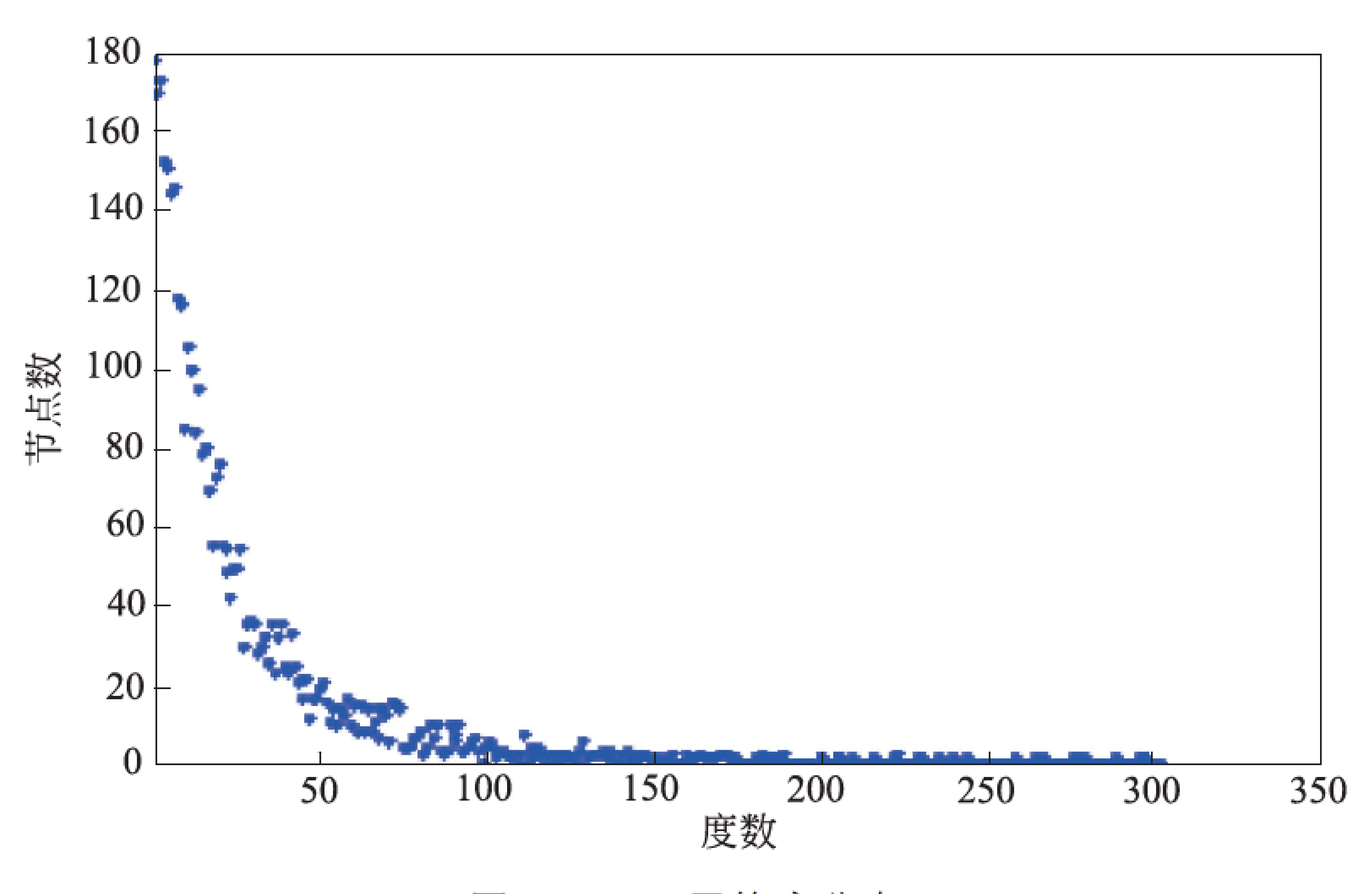


图 5-7　BA 网络度分布

与 BA 无标度网络的特性一样，本节模拟实验用到的 BA 网络的节点度分布呈幂律分布。在该网络上，模拟 ICSIS 模型传播过程，要与 5.2.5.1 节中的 WS 小世界网络上的传播过程进行对比，则控制变量、实验的条件设定与 5.2.5.1 节中的设定一致，如表 5-10 所示，信息 A、信息 B 的传播率分别设为 0.5 与 0.6，两种类型信息的恢复率都为 0.1，信息 A 与信息 B 的取代率分别为 0.4 与 0.6，吸引力分别为 0.4 与 0.6。信息 A 在传播开始就已经在网络上以多源方式传播，相隔 5 个迭代周期后，信息 B 开始在网络上传播。信息 B 的强度大于信息 A 的强度，如此设定是希望看到信息 B 能够抑制甚至覆盖信息 A 的传播，

5.2.5.1 节得到了所希望的结果，信息 B 不仅抑制了信息 A 的传播，而且完全覆盖了信息 A 的传播。同时可以看到，网络中存在稳态，由于网络稳定态分析中得到的平衡点与网络节点度分布有关或是与网络图的邻接矩阵的特征根有关，而小世界网络与无标度网络节点的度分布不同，那么 BA 网络的传播模拟结果会是怎样的呢？通过前面提到的条件设定，在节点数为 4000 时的 BA 网络上传播模拟的结果如图 5-8 所示。

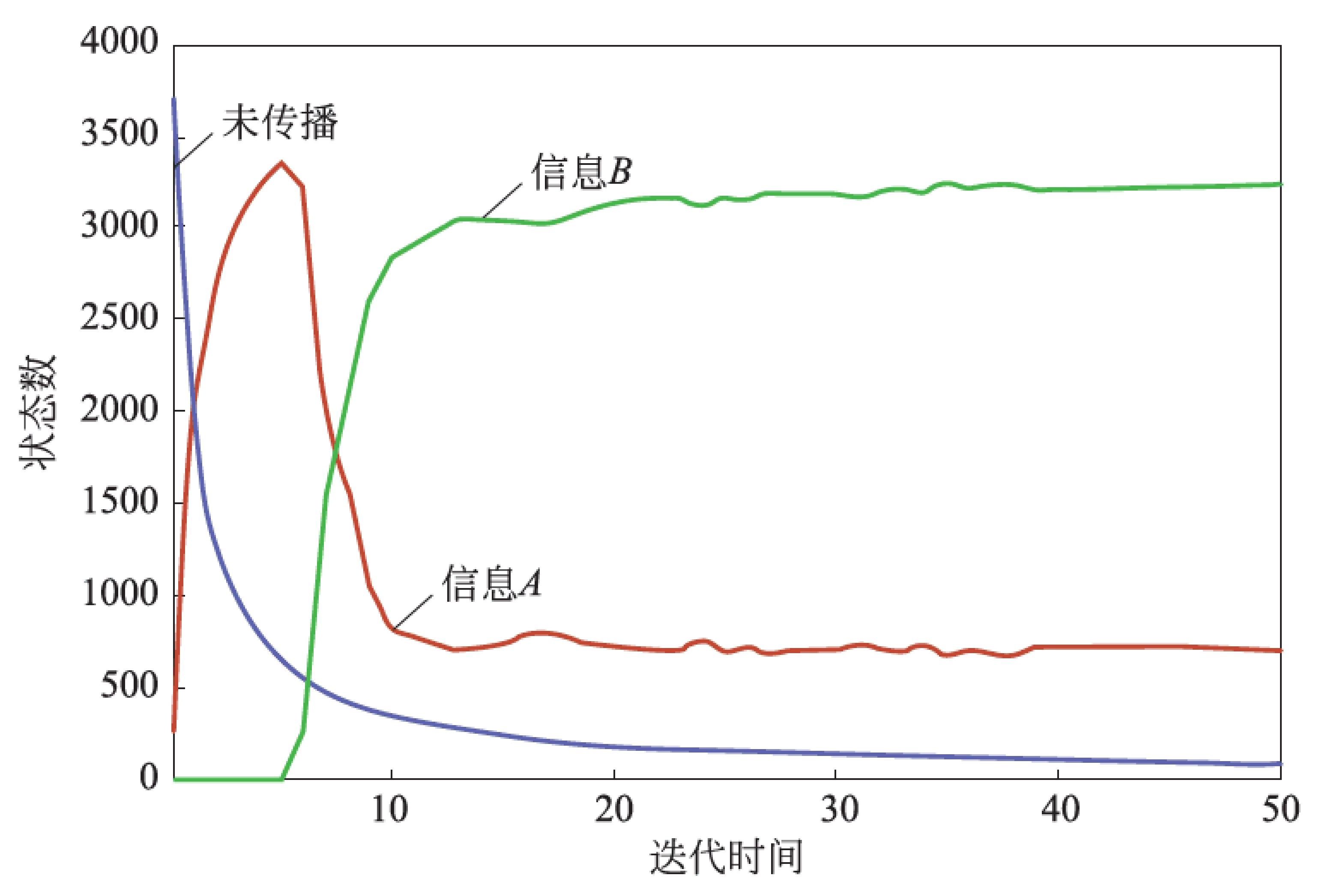


图 5-8　BA 网络信息传播过程网络状态变化

信息 B 传播前网络上只有信息 A 独立传播，由于缺少竞争，信息 A 逐渐传播扩散到网络的大部分地方；当信息 B 开始传播时，信息 A 的传播便开始受到抑制，信息 B 的传播逐渐覆盖信息 A 的传播，扩散至网络的大部分范围。最终，信息的传播达到一个动态平衡的稳定态。

5.2.5.3　网络拓扑结构对传播过程影响的结论

在 ICSIS 模型中，网络拓扑结构在其中扮演着非常重要的角色，无论是从宏观角度的节点度分布，还是在概率模型中的邻接矩阵，很容易理解度分布其实就是邻接矩阵对应的列的叠加，邻接矩阵 A 与节点 i 的度 k_i 的函数关系有

$$k_i = a_{ii}^{(2)}$$

邻接矩阵二次幂 A^2 的对角元素 $a_{ii}^{(2)}$ 就是节点 i 的邻边数。所以不难理解网络拓扑结构在 ICSIS 模型中的重要性。由于小世界网络的度分布呈正态分布，而无标度网络度分布呈幂律分布，二者的网络结构存在很大的区别。小世界网络中每个节点间的平均最短路径很短，正如社会学中的六度空间理论所描述的，小世界网络表现出了社会网络的小世界效应，小世界网络中节点的度数分布均匀，而无标度网络中节点的度数分布极为不均，多数节点与其他节点的连接很少，这使得网络中节点的平均路径长度较长，从而使得信息在传播过程中渗透得较慢。如上所述，在 WS 小世界网络和 BA 无标度网络的传播模拟实验结果中显示，信息 B 在 WS 小世界网络中传播对信息 A 的抑制效果较于在 BA 无标度网络中，不仅快而且强。在 WS 小世界网络中，信息 B 抑制信息 A 只用了 3 个迭代单位时间，而 BA 网络中，信息 B 抑制信息 A 的传播需要多于 5 个迭代单位时间；抑制强度上，WS 小世界网络中，信息 B 完全覆盖了信息 A 的传播。而在 BA 无标度网络中，信息 B 的传播只是大部分覆盖信息 A 的

传播，并未完全覆盖，分别占网络的 3/4 和 1/4。这样的结果很容易解释和理解，小世界网络中节点与节点间的连接平均且平均路径较短，因而信息 B 的影响会很快蔓延，同时，小世界网络中节点间的连接比较充分，因而信息 B 在覆盖信息 A 的传播时会更加彻底；而对于 BA 无标度网络，节点间距离较远，且节点连接不均匀。同时，节点间的连接不够充分，因此，信息 B 在 BA 网络中传播时对信息 A 的影响扩散得较慢，对于有些节点的影响可以忽略不计，因此二者在 BA 网络上可以共存。

5.2.6　信息发布时间点对竞争性信息传播的影响

本节主要通过对信息发布时间点的不同来探究后信息发布对传播过程的影响。信息发布时间点对两种信息传播过程的影响主要来自模型中起始 p_{I_A}，p_{I_B} 二者的初始值。信息 B 作为晚发布的信息，其发布的时间决定了信息 A 已经在网络中的传播时间，这样，当 B 刚开始传播时，网络中各节点的状态分布密度是不同的，信息 B 发布得越早，那么信息 A 的传播范围越小；信息 B 发布得越晚，则信息 A 的传播范围越大。这样，当两种信息共同在网络中传播的起始时间，也就是说，信息 B 在网络中开始传播时，网络中节点的传播状态分布会因其相对信息 A 滞后发布程度的不同而有很大不同。本节用 ICSIS 模型来研究这种现象。模拟实验的传播参数设定如表 5-10 所示，信息 A 的传播率为 0.5，信息 B 的传播率为 0.6，信息 A 与信息 B 的恢复率都为 0.1，信息 A 的取代率为 0.4，信息 B 的取代率稍大，为 0.6；信息 A 的吸引力为 0.4，而信息 B 的吸引力为 0.6。在本次模拟实验中，其他参数设定如表 5-12 所示。

表 5-12　信息发布时间点不同模拟条件设定

参　数	设定值
网络拓扑结构	BA 无标度网络
网络起始节点数	50
网络每次信息节点数	40
网络节点总数	4000
信息 B 发布相对时间点	0,2,5,10,20

按照表 5-10 和表 5-12 的条件参数设定，分别对信息 B 的发布时间相对时间点 0,2,5,10,20 进行模拟传播过程，总体网络分布迭代情况如图 5-9 所示。

信息 B 的发布时间点选择为信息 A 开始传播后的 0 个迭代单位时间、2 个迭代单位时间、5 个迭代单位时间、10 个迭代单位时间和 20 个迭代单位时间，分别对应信息 A 的起始传播、传播覆盖一半网络、传播覆盖大半网络、传播覆盖大部分网络、几乎覆盖整个网络几个时间节点。图 5-9 是 5 个时间测试点的整个网络传播过程总图，为了更清晰地了解信息 A 的被抑制情况，图 5-10 给出了不同时间间隔的控制信息发布时间点下信息 A 的传播情况。

对于信息 A 的传播过程，若信息 B 发布时间点越晚，则 A 的传播终态将覆盖更大的范围，被抑制效果较差。图 5-11 给出不同信息 B 发布时间点下信息 B 的传播过程，与图 5-10 一起讨论，将更清晰地表明不同的信息 B 发布时间点对传播过程的影响。

对于不同的信息 B 的发布节点，在 ICSIS 模型中，初始值 p_{I_A} 从 0～1 随间隔时间的增加而增大，而 p_{I_B} 一直都是 0。信息 B 发布时间点越晚，则信息 A 传播覆盖的范围越广，那

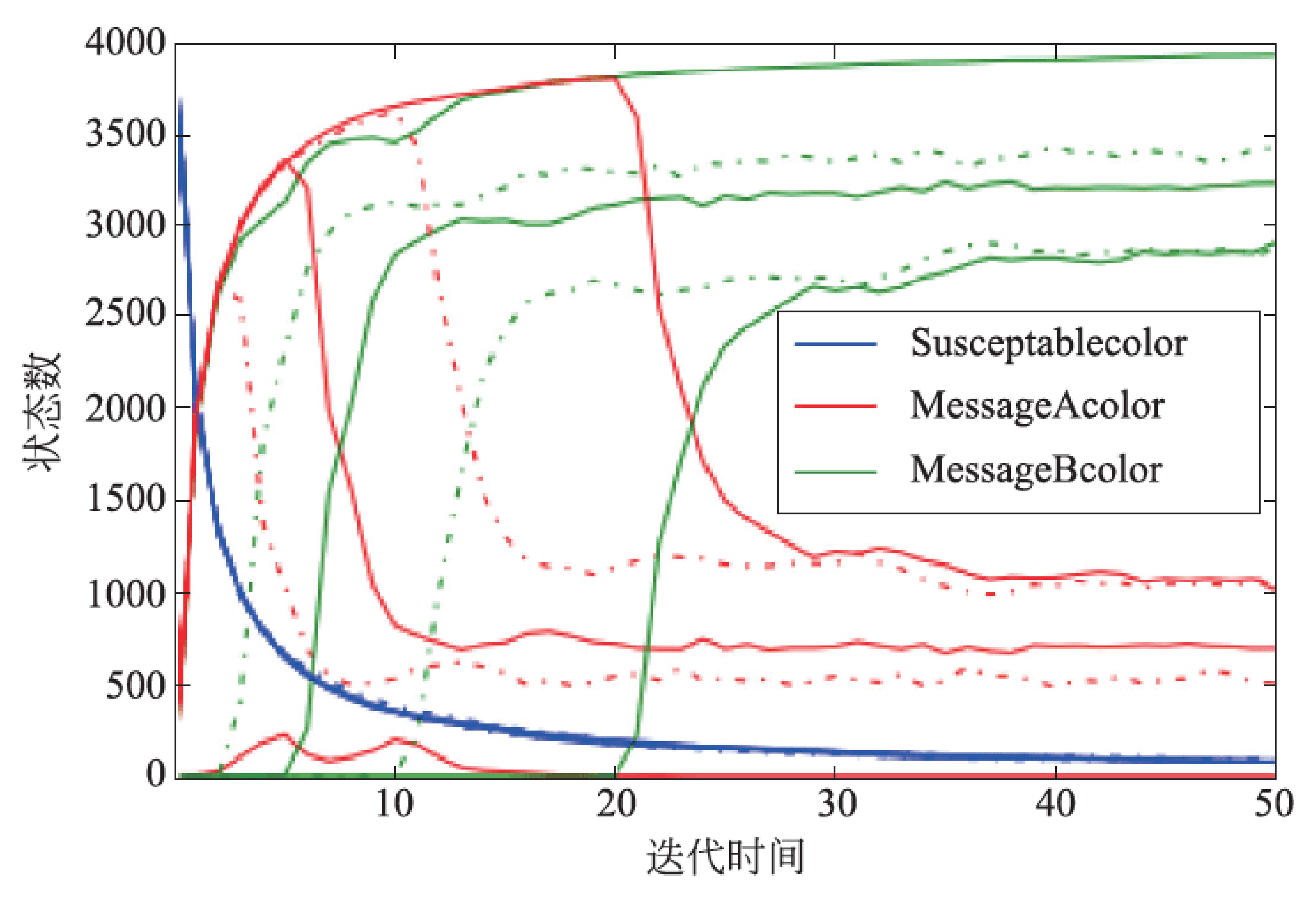


图 5-9　信息 *B* 不同发布时间点网络节点状态总图

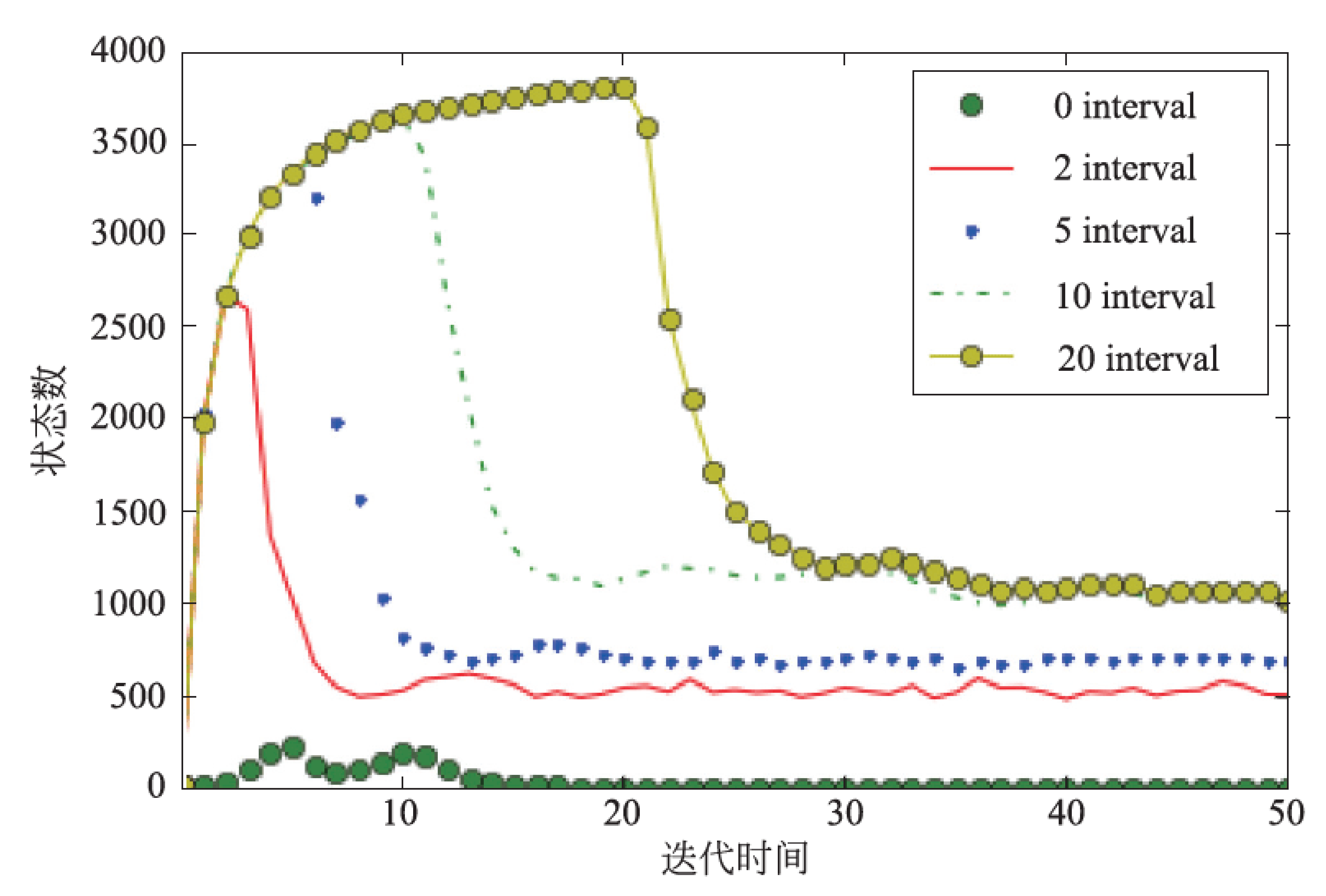


图 5-10　信息 *B* 不同发布时间点时信息 *A* 的传播情况

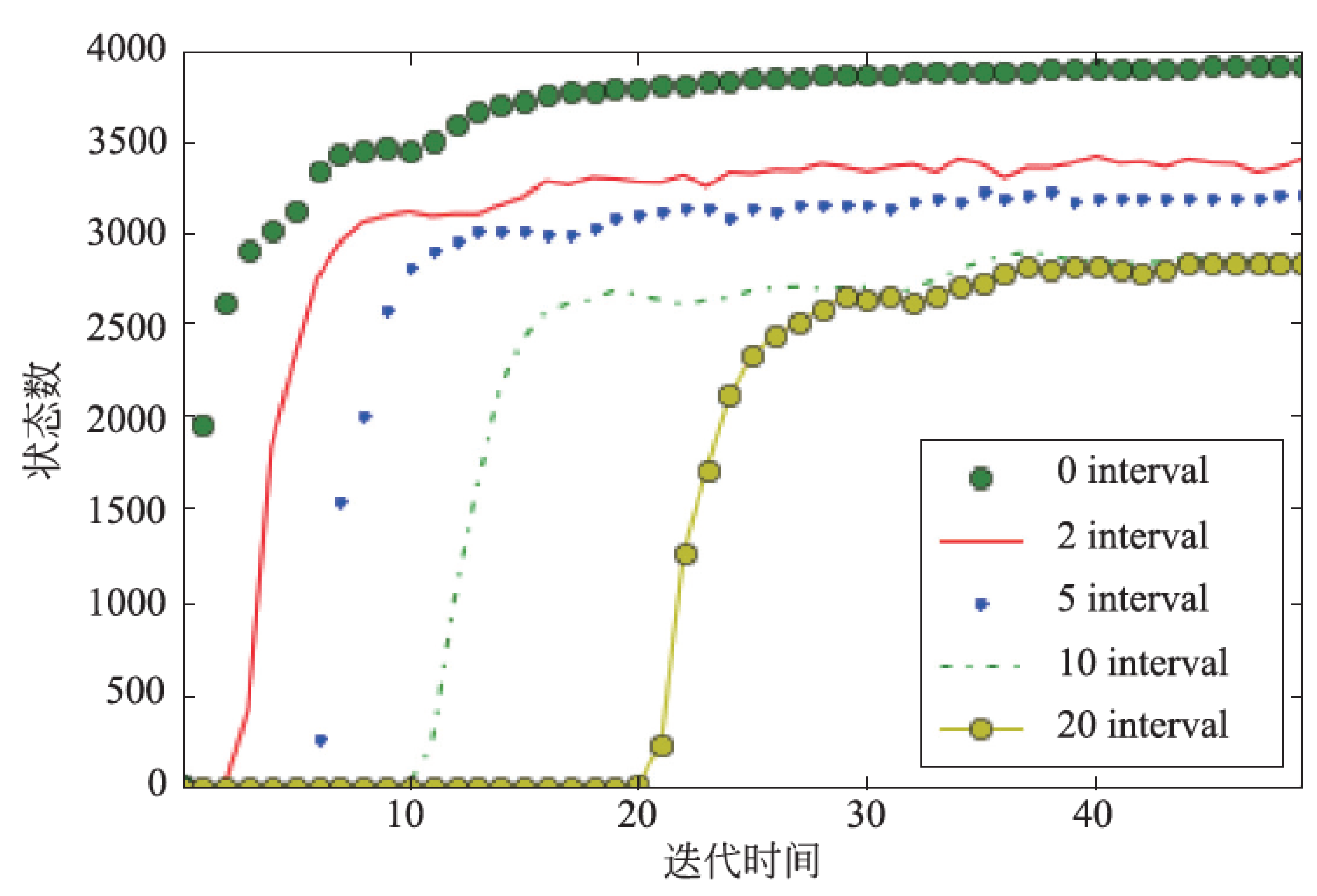


图 5-11　信息 *B* 不同发布时间点时信息 *B* 的传播情况

么强度更大的信息 B 需要更长的时间去抑制信息 A 的传播。同时,抑制效果不如相对早的时间点发布时的情况,越晚发布,则最终网络中信息 A 的传播范围残留将会相对较大。不过,当时间点超过某个点后,将不影响传播的结果,从图 5-11 可知,当 10 与 20 个迭代单位时间点发布时,抑制曲线相对平行,最终的状态线几乎重合,也就是说,当信息 A 传播扩大到几乎整个网络的时间点以后,时间点的选择对抑制结果相差不大。其中,信息 A 传播覆盖范围的起始值与信息 B 的发布时间间隔的关系如图 5-12 所示。

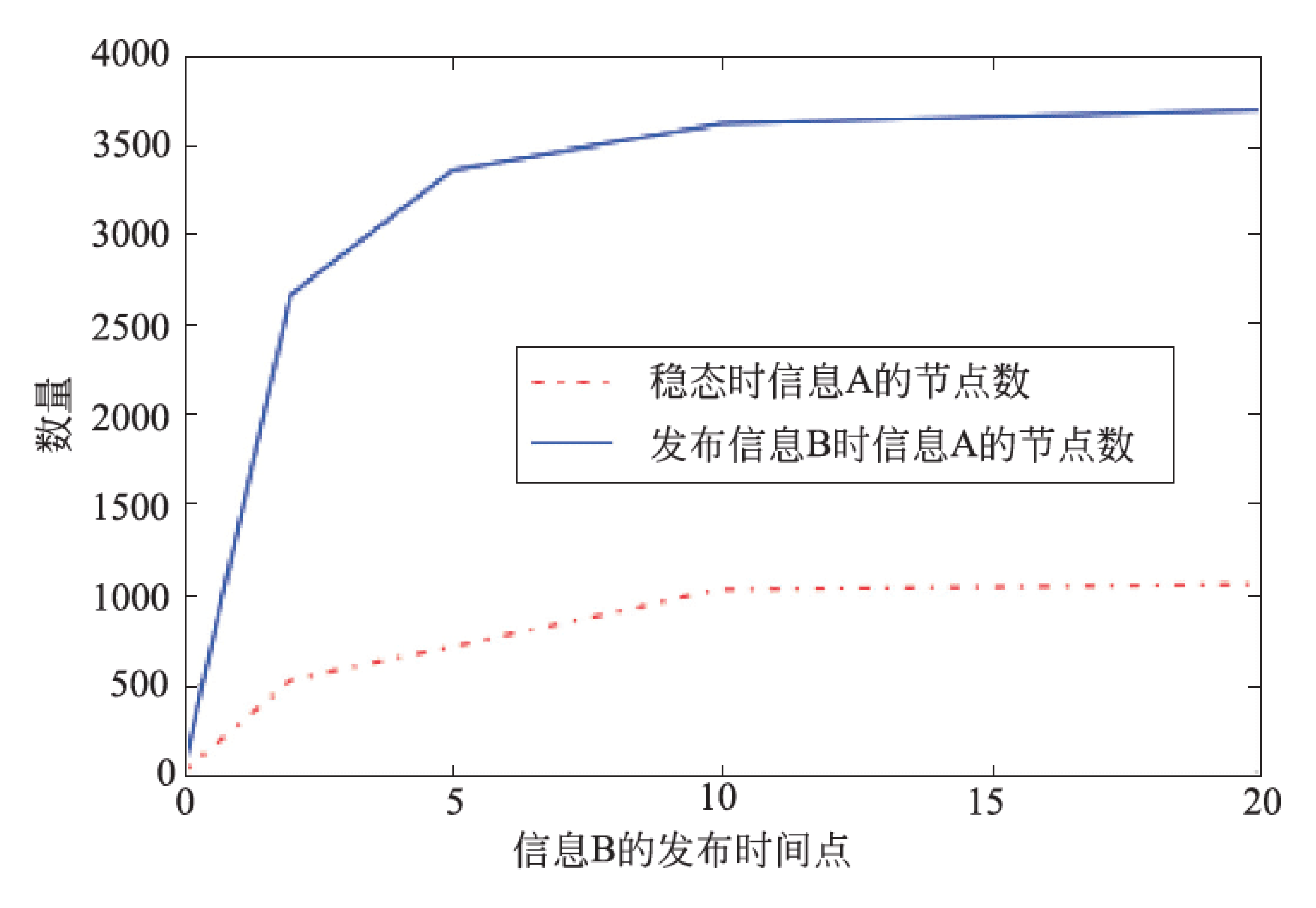


图 5-12 信息 A 的传播范围与信息 B 发布时间点的关系

信息 A 的起始传播范围随信息 B 发布时间间隔的增加呈一个类似对数函数的增长态势,当发布时间越长时,信息 A 的范围增加越发平稳。随着时间间隔的增大,抑制效果越来越差。因此,如果想通过信息 B 的发布抑制信息 A 的传播,则需要尽早发布。当一个信息在网络上开始蔓延时,对其控制的措施要及时,等蔓延的范围广泛时,管控效果将会降低。

5.2.7 网络节点特征对竞争性信息传播的影响

本节将探究网络节点特征对信息传播过程的影响,网络中节点的特征包括在线频率、节点度数等,本节对网络中节点的在线频率与节点度这些特征对传播的影响进行探究,通过对信息 B 的发布节点进行控制,探究不同在线频率的信息 B 起始传播节点与不同节点度数的信息 B 起始传播节点对最终网络状态的影响。

5.2.7.1 基于不同在线频率的信息 *B* 起始传播节点探究

用户在线频率 p_f 表示节点的在线的概率,p_f 越大,那么该节点相对越活跃,在线的时间也会越长。简单地理解就是,该用户花费很多时间刷微博等社交媒介,p_f 越小,则该节点相对不活跃,在线时间较短,对社交媒介的兴趣可能不大或者个人自律性好,严格控制自己的社交网络在线时间。通过模拟实验探究信息 B 起始传播节点的在线频率与最终网络状态的关系,本模拟实验的传播系数设置如表 5-10 所示,信息 A 和信息 B 的传播率分别为 0.5 和 0.6,信息 A 与信息 B 的恢复率都为 0.1,信息 A 与信息 B 的取代率分别为 0.4 与 0.6,信息 A 与信息 B 的吸引力分别为 0.4 与 0.6。本次模拟实验的其他条件设定如表 5-13 所示。

表 5-13　节点频率模拟实验参数设定

参　数	设定值
网络类型	BA 无标度网络
网络起始节点数	50
网络每次信息节点数	40
网络节点总数	4000
信息 B 发布与信息 A 发布时间间隔	5
节点频率范围	0～2,4～6,8～10

使用上面的模拟实验条件进行传播模拟实验,得到网络中节点所有状态变化的总图,如图 5-13 所示,在稳定态,信息 B 起始节点的在线频率越高,则信息 B 对信息 A 传播的抑制效果越好。

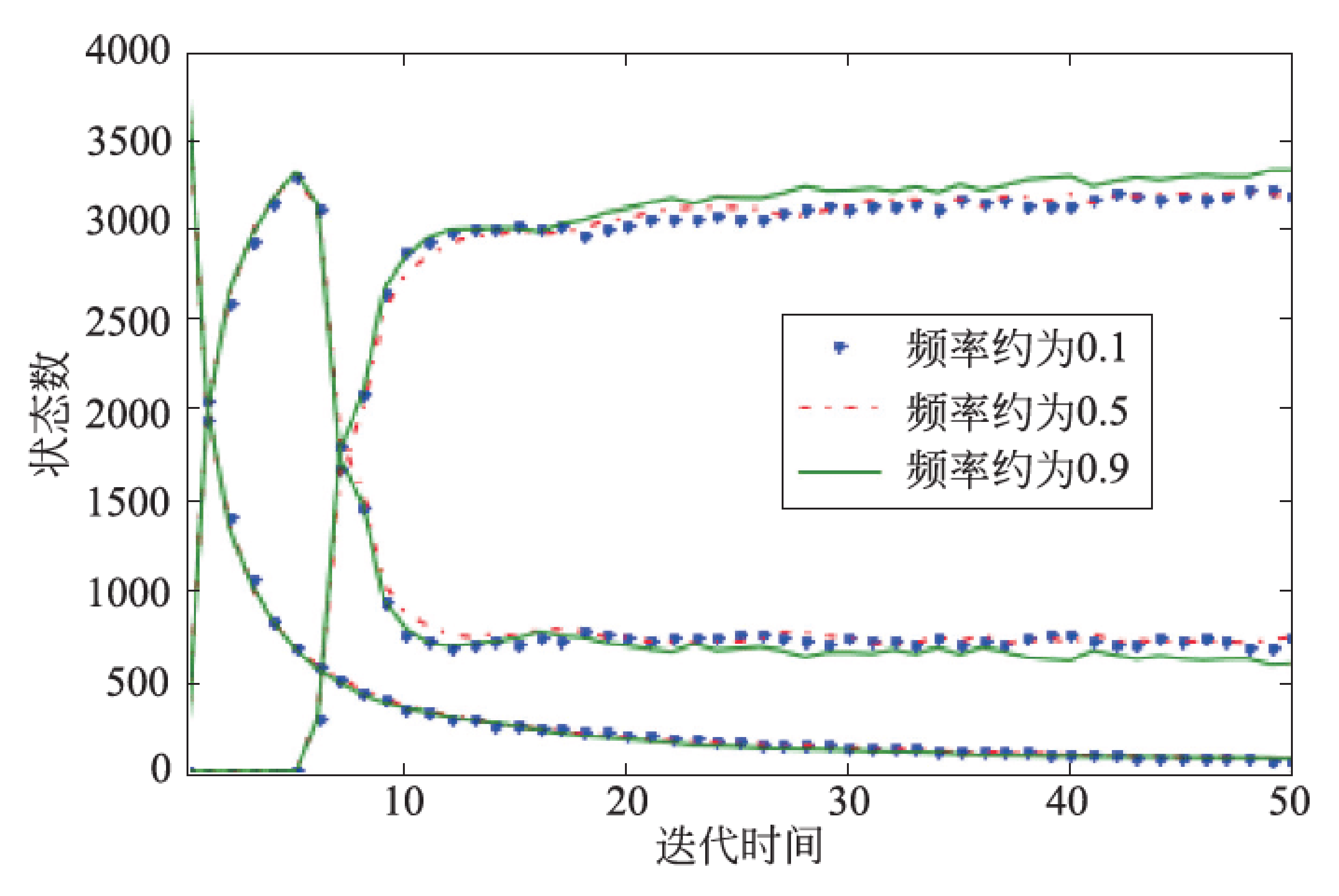


图 5-13　信息 B 起始节点频率与网络终状态关系总图

信息 B 的起始频率不同,则网络达到稳态时信息 B 最终在网络中的覆盖范围略有差异,起始节点频率与信息 A 稳态时覆盖范围和信息 B 稳态时覆盖范围的关系如图 5-14 所示,随着在线频率的提升,抑制效果略有提高。

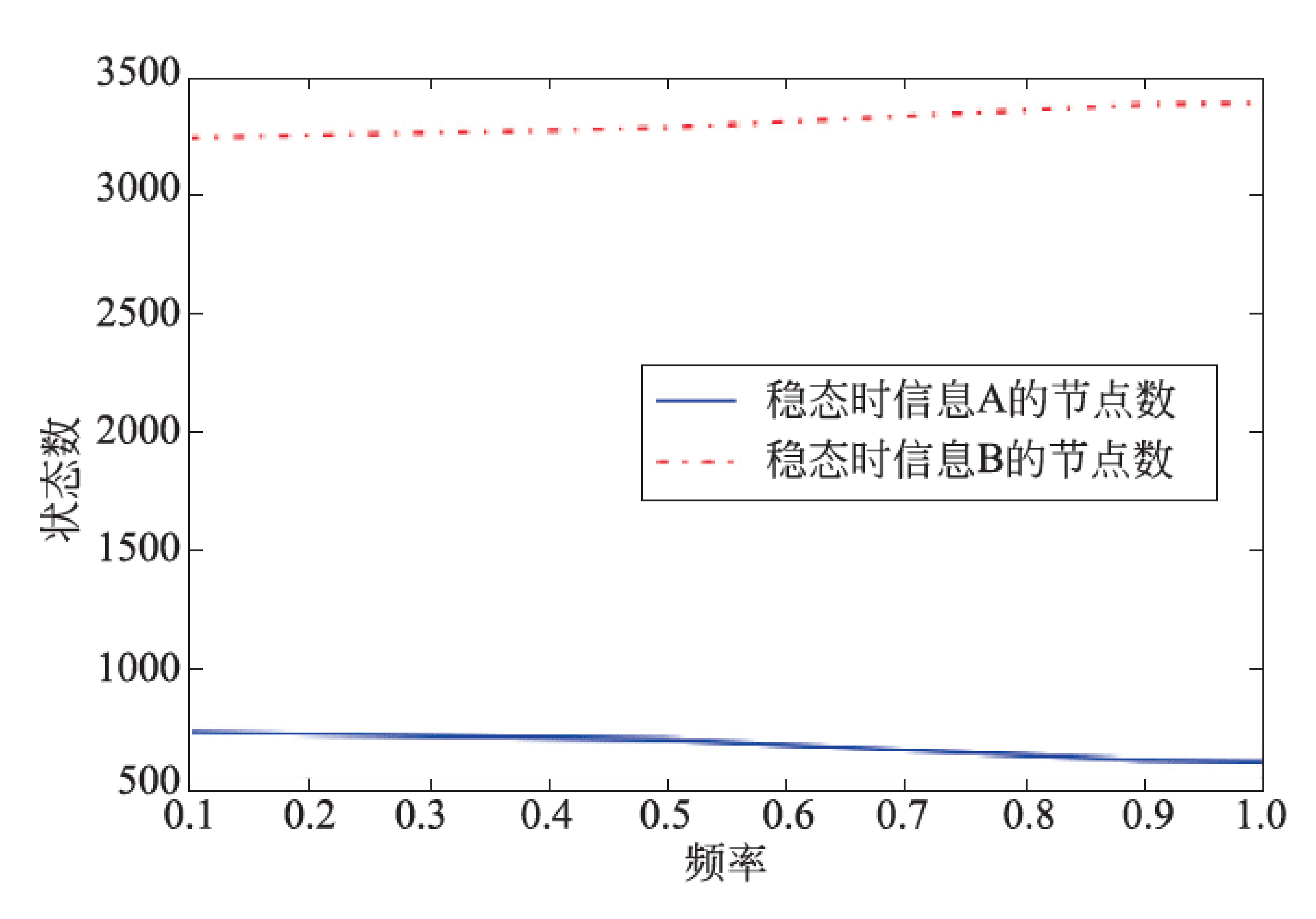


图 5-14　信息 A 和信息 B 稳态覆盖范围与频率关系

当频率越高时，信息 B 对信息 A 的抑制效果较好。不过，抑制效果的优势不是很明显，比较微弱，对于频率较高的节点，其节点度数未必最佳，这就导致了该点未必是最佳选择，不过，单方面讨论的话，当节点的在线频率越高时，即该点在线时长较长，那么该点便存在更大的概率传播消息，因此，当信息 B 的初始传播节点的在线频率越高，则信息 B 对信息 A 的抑制效果将会更好。

5.2.7.2　基于不同度数的信息 *B* 起始传播节点探究

节点的度数在实际中的反映便是用户的好友数量，对于不同的复杂网络结构，网络中都有不同的节点度分布，本小节将对不同度数的信息 B 的起始传播节点与最终网络状态的关系进行探究，通过模拟实验得到模拟情况下二者的关系，本次模拟实验传播过程的参数与表 5-10 所示一致，信息 A 和信息 B 的传播率分别为 0.5 和 0.6，信息 A 与信息 B 的恢复率都为 0.1，信息 A 与信息 B 的取代率分别为 0.4 与 0.6，信息 A 与信息 B 的吸引力分别为 0.4 与 0.6。本次模拟实验的其他实验条件设定如表 5-14 所示。

表 5-14　节点度数模拟实验参数设定

参　数	设定值
网络节点总数	4000
重连边数	100
重连概率	0.6
信息 B 发布与信息 A 发布时间间隔	5
节点度数范围	80～90，90～100，100～110

根据上述的实验条件设定进行信息 B 发布节点的度数对信息传播的影响的分析，对于该实验设定的 WS 小世界网络，其网络的节点度分布如图 5-15 所示。

本次实验的节点度数分布呈现以 100 为中心轴的类正态分布，因此，模拟实验时，选择 80～90，90～100，100～110 度数区间进行模拟实验，得到信息 B 的发布节点的度数与传播

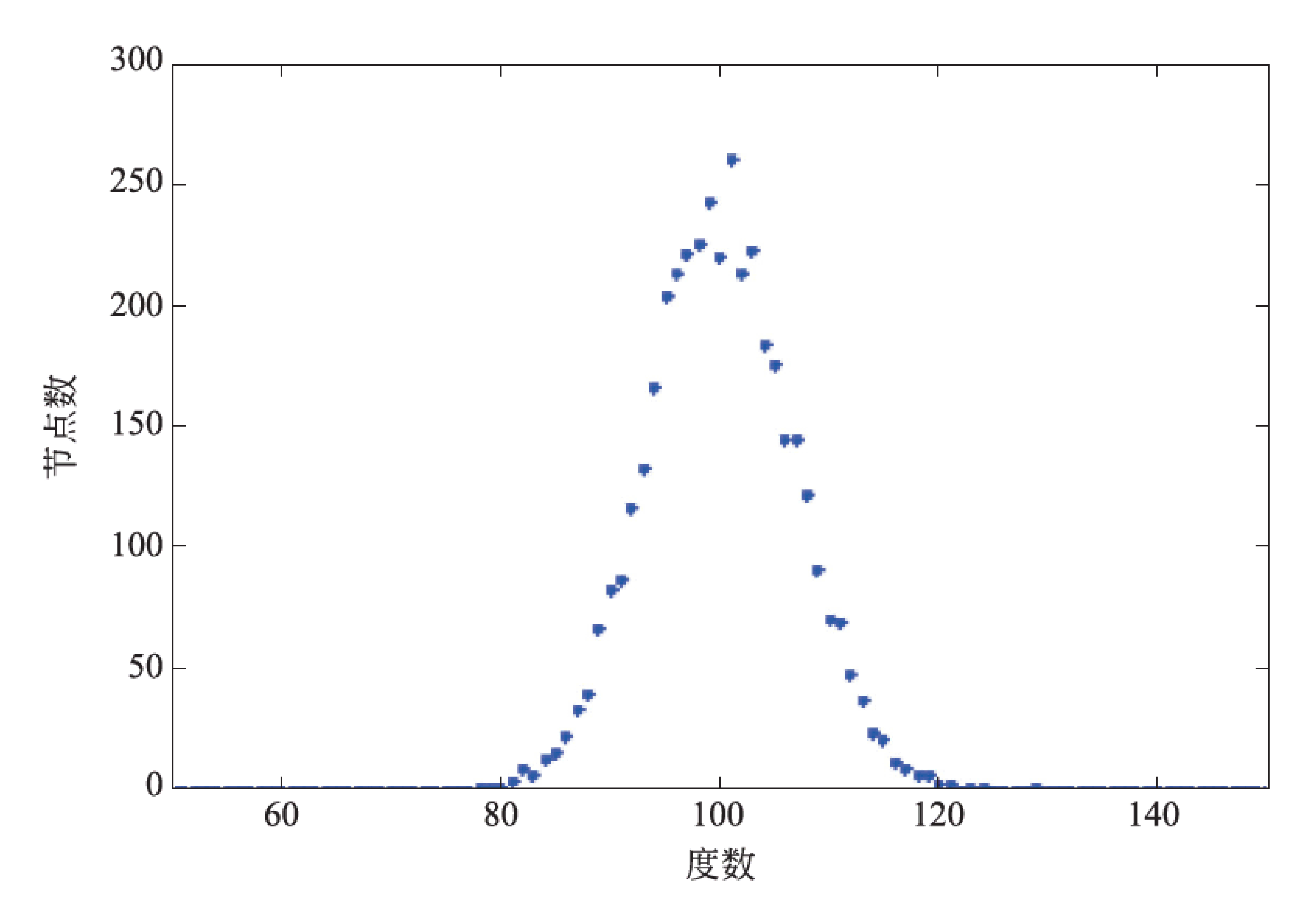


图 5-15　节点度数模拟实验网络度分布

过程中的网络状态的关系总图，如图 5-16 所示。

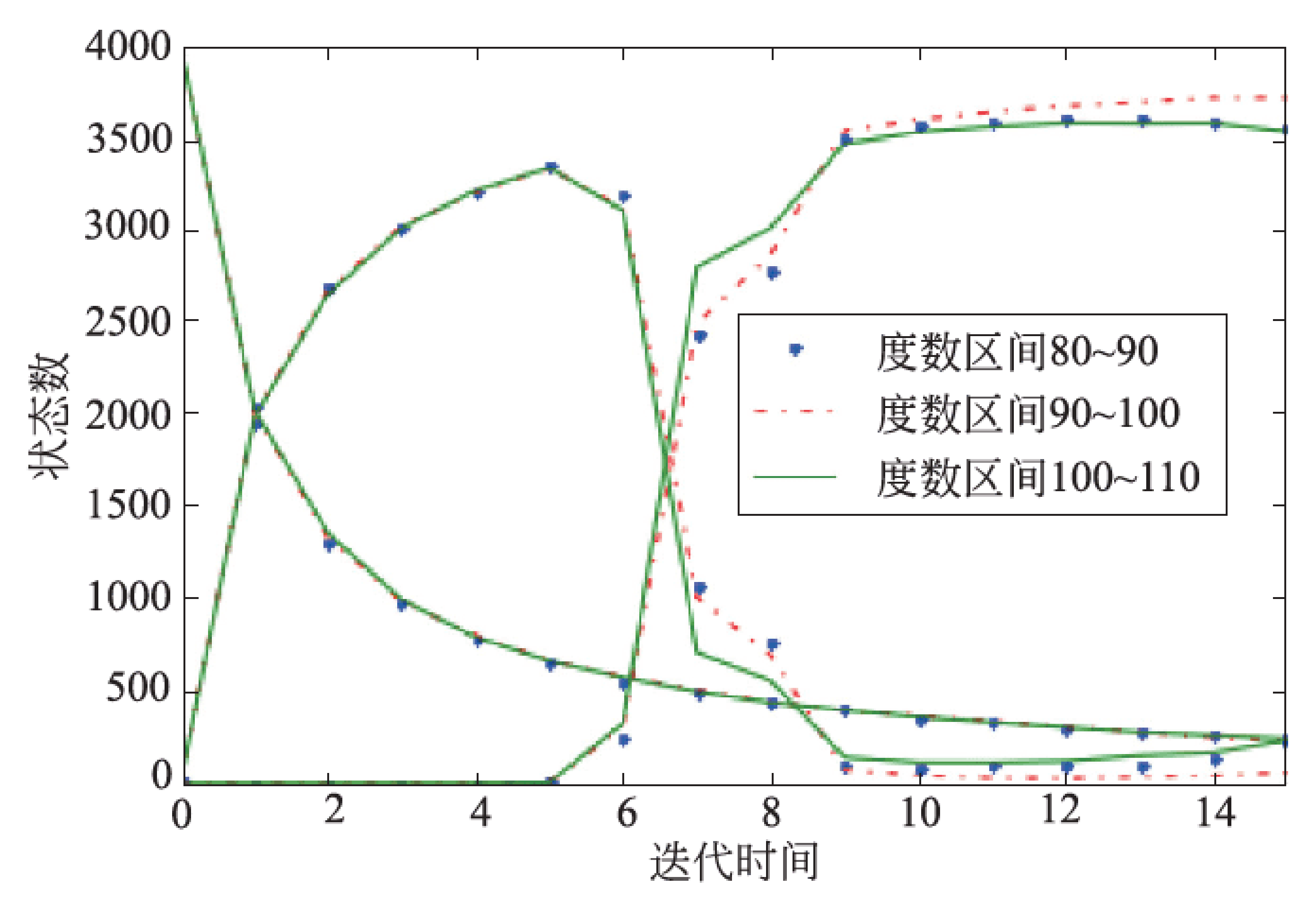


图 5-16　节点度数与传播过程的关系

调控信息发布节点的度数越高，则对于抑制已传播信息的传播的速度越快，即调控信息的覆盖效率更高，而在最终的稳定态，在网络中该度数节点的密度越大，则调控信息的覆盖范围越广，即已传播信息的被抑制效果越好。

5.2.8　竞争信息传播演化过程分析

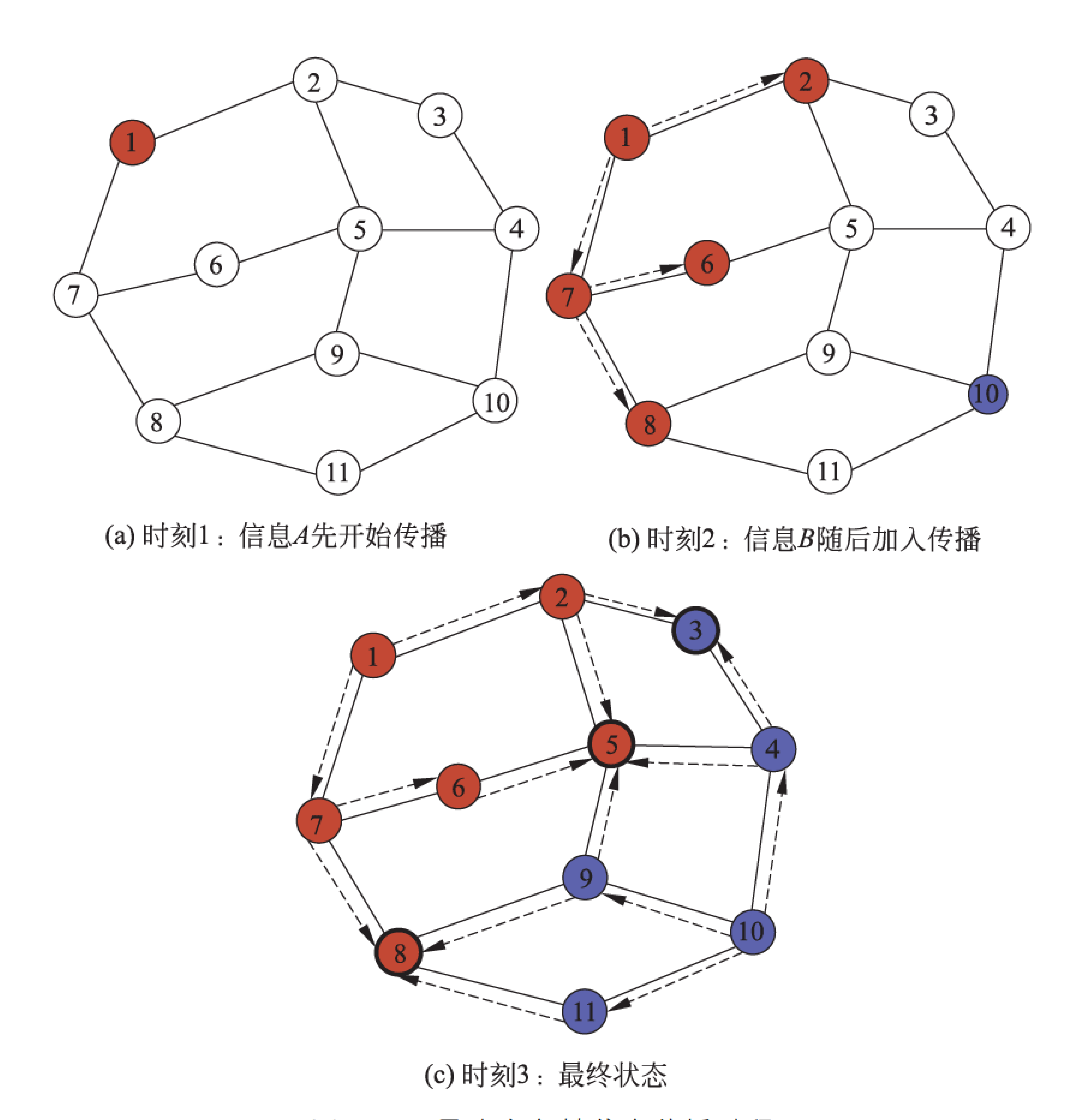
在线社交网络应用为广大的互联网用户提供了一条便捷的与其他网络用户沟通交流的通道，同时也为网络上产品的推广、广告的推送搭建了一个更为宽广的平台。相关领域内的两类相似产品或者广告在社交网络上进行推广，很大程度上会产生相互竞争的现象。网络个体对某一类产品或者广告的偏好受用户对其接受程度以及网络上从众心理的共同影响。在线社交网络上不单单存在信息同步竞争的现象，更为广泛存在的是异步竞争传播这一行为。某一类的信息产生的时间往往有所差异，但同时又是针对同一目标群体。网络上的广告推广、商品营销更多的是以异步竞争传播形式而存在的。

本节利用社交博弈模型分别对社交网络中竞争信息的异步方式和同步方式的传播过程进行演化分析。

5.2.8.1　竞争信息异步传播演化过程分析

异步信息传播是由社交网络中的用户针对在某一时间段内，从某一个特定主题所产生的多个相关信息，先后在网络上开始传播并相互竞争影响，用户与其邻居进行交流和分享信息或者意见。用户自身对信息接收程度以及个人偏好同样影响传播结果，因此也可以看作是一个考虑用户个体声誉因素影响，基于信息交互的博弈过程。异步信息传播过程如图 5-17 所示。红色和蓝色分别代表了网络上的两种先后加入竞争的信息 A 和信息 B。

在图 5-17 所示的有向网络描述的社交网络中，圆圈表示用户，圆圈之间的连线表示用户间的社交关系。在传播初始阶段，竞争性信息 A 进入社交网络，率先影响网络用户。当红色信息 A 已经传播一段时间后，影响了一批用户后，蓝色信息 B 才进入网络，加入信息竞争当中。网络用户对于是否接纳该信息不仅出于信息对自身所产生的吸引力以及带来的收



(a) 时刻1：信息A先开始传播　　(b) 时刻2：信息B随后加入传播

(c) 时刻3：最终状态

图 5-17　异步竞争性信息传播过程

益的考量，而且还受到周围网络用户，即用户的社交好友，对于该信息接纳与否的行为影响。信息 B 进入网络开始传播时，网络中已经接收信息 A 的用户数对信息的竞争选择可能会造成一定影响。如果信息 A 尚未广泛传播开来，此时信息 B 相对较容易与信息 A 竞争。但是，倘若信息 A 已经传播了相当一段时间，网络中影响对信息 A 的接收度已经很高，此时信息 B 想要在竞争中取得更好的传播效果，就必须额外地多为用户提供一些收益来使用户改变他们的信息选择。信息所具有的价值越高，就越能吸引网络用户的关注。随着信息竞争演化传播的进行，网络上的信息传播最终达到一个相对稳定的均衡状态，如最终状态图 5-17(c)所示。两类竞争性信息在社交网络中按一定比例共存，用户的信息选择以及周围用户如何选择信息的行为，对信息最终的传播结果均会产生一定的影响。

社交演化博弈模型为解决竞争性信息传播问题提供了一个很好的方法思路。在社交网络博弈模型中，每个博弈个体不仅追求策略信息所带来的短期效用，而且还关注关系结构调整所带来的长期声誉影响。相应地，模型中也建立两种更新机制，分别对应于获得更大效用的策略更新机制和更高声誉的社交关系更新。

1. 模型设置

竞争性信息异步传播模型采用基于第 4 章中提出的社交演化博弈(SEG)模型。采用无权有向的均匀度随机网络 $G=(V,E;I,U,R)$表示。V 表示个体集合；E 为关系集合；I 为个

体间信息行为集合；U 代表效用函数集；R 为声誉函数集。用户个体可以采用合作 $C(s=[1,0]^{\mathrm{T}})$ 或背叛 $D(s=[0,1]^{\mathrm{T}})$ 两种策略。个体 i 有两类邻居：入邻居 N_i^I 与出邻居 N_i^O，对应的入度为 d_i^I 和出度为 d_i^O。

考虑到在线社交网络具有网络效应，网络中的用户个体不仅追求信息影响力所带来的收益，而且还关注在转发传播信息时对自身声誉所造成的影响。协调博弈可以很好地描述具有网络效应的多种因素影响下信息竞争的传播现象。根据协调博弈，可以做出以下两种假设：①某一类信息在社交网络上传播的范围越广，影响力越大，越能吸引网络用户关注转发这一类信息；②网络用户一旦已经接纳某一类信息，那么用户不会轻易改变已采纳的信息，转而转发相似信息。一方面，用户个体为了某一共同目标协调各自行为以获得更大收益，并且用户获得的收益随着协调行为中合作人数的增加而增大；另一方面，一旦用户选择某一信息策略，那么他背叛已有信息策略的代价将非常高。

协调博弈是一个有多个纳什均衡的博弈，博弈个体试图加强相互信任以期达到某个均衡点，而避免出现非均衡点的结果。网络用户个体坚持某一共同策略，或者共同选择背叛，转而选择新信息策略均可达到纯纳什均衡点。相应地，两个均衡点分别成为收益占优均衡和风险占优均衡。在均衡点上个体的收益相同。此时，个体努力通过合作避免出现非均衡的结果，博弈个体关心的是博弈结果出现在均衡点上，而不关心出现在哪一个均衡点上。在这个意义上，个体之间的目标是一致的，在这样的博弈中个体的利益相关但不存在冲突。而在更为广义的协调博弈中，存在多个均衡，在任何一个均衡点上所有个体的结果都好于非均衡点，但是个体在不同均衡点上的收益是有差异的。在这样的协调博弈中，每个博弈个体不仅关心避免博弈结果在非均衡点上实现，这是所有博弈个体的共同目标，在这个意义上他们是合作的。同时，每个人希望博弈在其所偏好的均衡点上实现，在这个意义上他们是竞争的或者是冲突的。协调博弈中，信息收益以及用户行为期望和对风险和收益的预期，对竞争性信息传播的过程产生不确定性。

使用一个 2×2 矩阵 $\begin{pmatrix} a & c \\ d & b \end{pmatrix}$ 来描述协调博弈，其中，$a>d$，$b>c$。协调博弈中博弈个体可选择策略 A 或策略 B。对应该矩阵，共同选择策略 A 或者策略 B 均为纯纳什均衡点：a 为收益占优均衡，b 为风险占优均衡。效用是影响协调博弈均衡的一个因素，除此之外，博弈均衡还与用户行为以及对风险和收益的预期相关。a 和 b 两个纯纳什均衡点是用户个体在竞争性信息传播博弈过程中所期望达到的状态。而当博弈双方选择不同策略信息，即 c 或者 d 时，协调博弈则处于非纳什均衡点，对于用户而言应尽量避免达到这种状态。因此，协调博弈可以很好地描述竞争性信息传播中的特性。

本节假设个体是非理性且自私的，博弈交互时只关注效用因素。个体因为短视只能获得其邻居节点与邻居节点的邻居节点的信息。每个个体的效用是其参与所有博弈所获得的收益之和，即

$$u_i = \sum_{j \in N_i} s_i^{\mathrm{T}} M s_j$$

其中，N_i 是个体 i 所有参与博弈的选择邻居集合，s_i 和 s_j 分别为博弈双方个体 i 和 j 的策略，M 是所采用的协调博弈的 2×2 收益矩阵。对于异步竞争传播，采用一般协调博弈（general coordination games），其收益矩阵表示为 $M=\begin{pmatrix} 1 & c \\ c & b \end{pmatrix}$。当博弈双方相互合作时，双方均获得

收益为 1；当博弈双方均选择背叛策略时，获得的收益为 $0\leqslant b\leqslant 2$；而当一方合作，另一方背叛时，双方获得的收益为 $0\leqslant c\leqslant 1$。当双方背叛收益 $0\leqslant b<1$ 时，博弈双方通过选择合作策略才能获得更大收益，这时处于收益占优的博弈状态；而当 $1<b\leqslant 2$ 时，共同选择背叛策略对博弈双方的诱惑力更大，这时处于风险占优的博弈状态。

在异步更新机制中，在时刻 t 随机选中一个个体，依据$\frac{1}{1+W}$概率进行行为策略更新，否则，以 $1-1/(1+W)$进行社交关系更新。$W=\frac{\tau_e}{\tau_p}$ 为策略更新时间尺度τ_e和社交关系更新时间尺度τ_p的比值。每个个体的声誉$R_i(t)$按照以下公式进行更新。

$$\begin{cases}R_i(t)=\sigma R_i(t-1)+\Delta R_i(t)\\ \Delta R_i(t)=\sin\left(\frac{\pi}{2}\frac{n_i^C}{d_i}\right)\end{cases}$$

其中，n_i^C表示个体 i 合作邻居的个数。

当个体 i 进行行为策略更新时，选择最大效用邻居，满足

$$j=\arg\max_{l\in N_i^O}\{u_l>u_i\}$$

并模仿其行为策略，模仿方式遵循 Fermi 规则：

$$w(s_i\leftarrow s_j)=\frac{1}{1+\exp[\beta(u_j-u_i)]}$$

个体 i 更新其社交关系时，找到声誉最差邻居，并以概率p_s切断与它的关系，满足

$$j=\arg\min_{l\in N_i^O}\{R_l(t)<R_i(t)\}$$

依据声誉最大原则，选择邻居并建立关系：

$$k=\arg\max_{N_i^I\cup_{l\in N_i^O}N_l^O(\{i\}\cup N_i^O)}\{R_l(t)>R_i(t)\}$$

否则，以概率 $1-p_r$随机选择。

2. 模型仿真分析

在模型仿真过程中，初始化时建立一个平均度数 $k=4$、个体总数 $N=10^3$ 的随机网络。演化初始，所有节点以相同概率随机选择合作策略或者背叛策略，即选择坚持原有信息或是转而采纳新信息，参数 *ratio* 表示初始网络中选择合作策略个体所占的比例。影响实验中一系列蒙特卡洛仿真的重要参数有 3 个，分别是p_r，W 和 *ratio*，这 3 个参数也将用来分析具有演化博弈特性的信息传播机制。p_r反映了博弈中个体对更好声誉环境的渴望程度。p_r取值范围为 0～1，步长为 0.1。当$p_r=0$ 时，个体完全不关注声誉因素，只是随机选择邻居节点；当$p_r=1$ 时，表示个体在进行邻居选择时完全遵循最大声誉原则。W 反映了个体关注关系更新的频率，关注关系更新频率随 W 的增大而增大，用户个体间的交流更加频繁。通常在在线社交网络中，大多数用户调整关注关系的频率并不高，因此，W 最大值设为 1。W 的取值范围为 0～1，步长为 0.1。当 $W=0$ 时，个体不存在关注关系更新，只有个体的行为策略更新；当 $W=1$ 时，个体关注关系更新和行为策略更新比例相等。*ratio* 反映了初始群体中已经选择合作策略，接纳原信息的百分率。*ratio* 取值范围为 0～1，步长为 0.1。*ratio*$=0$ 表示初始群体中尚未选择合作个体，没有个体选择接收原信息；*ratio*$=1$ 表示初始群体中个

体均选择合作，无人背叛而选择新信息。在协调博弈收益矩阵中两个变量参数 b 和 c，分别表示博弈双方均选择背叛策略和一方合作一方背叛时所获得的收益，而将双方均选择合作时所获得的收益设为单位 1。b 的取值范围为 0～2，步长为 0.04。当 $0<b<1$ 时，博弈处于收益占优纳什均衡；当 $1<b<2$ 时，博弈处于风险占优纳什均衡。c 的取值范围为 0～1，步长为 0.1。当 $c=0$ 时，表示博弈双方一旦选择不同策略信息，双方均无法获得收益；当 $c=1$ 时，表示当博弈双方只要其中一人选择背叛，他们也将获得较高的收益。c 值越大，对博弈中个体选择不同行为策略的刺激越大。

在模型仿真过程中，采用合作者比例 f_c 作为个体的信息传播行为的协同演化过程的评价指标。f_c 反映网络中选择接纳原信息的百分比，f_c 越大，表示原信息在竞争中占据主导地位，f_c 越小，表示后加入竞争的新信息更能赢得用户关注。当网络中合作者比例 f_c 在一定时间内波动非常小时，则认为网络的演化达到相对稳定状态。如果社交演化博弈无法达到相对稳定状态，则演化的仿真时间将执行相对较长的时间，比如 2×10^5 个周期。当进入相对稳定状态或者执行 2×10^5 周期之后，再演化 10^3 周期，f_c 即为这 10^3 个周期内合作者比例的平均值。为了减小随机因素的影响，模型仿真结果取 20 次独立仿真的平均结果。社交演化博弈模型的其他参数设置如下：策略模仿噪声 $\beta=0.01$，切断社交关系连边概率 $p_s=0.01$，声誉的记忆衰减因子 $\sigma=1$。

图 5-18 显示了在不同的 c 值下，f_c 作为 (p_r,b) 函数时受这两个参数的影响情况。首先分析关注关系更新频率 W 对博弈演化的影响。从该图中可以看出，当 c 确定时，W 接近 0，在 $b<1.2$ 的区域存在非常高的合作率。相反，在 $b>1.2$ 的区域几乎没有合作现象产生。当关注关系更新频率较低时，合作者比例受信息收益影响较大，当新信息诱惑足够大时，网络中个体开始放弃原有信息转而采用新信息，个体的效用将决定网络中合作者能否生存。当 W 增大时，b 值对合作者比例 f_c 的影响开始减小，f_c 能够维持在一个相对较高的状态。关注关系更新频率的增大可以提高网络中合作者的比例，用户频繁地调整关注关系，使得用户更容易做出转发、评论等信息交互行为，在考虑声誉环境条件下，用户更加谨慎地进行信息策略的选择，不会轻易改变已选择信息。当新信息的收益诱惑对个体策略选择影响不够大时，个体仍然会坚持原信息。

同样地，关注收益矩阵中 b 和 c 对合作者比例的影响。由于博弈双方均坚持原信息收益为 1，均选择背叛策略而选择新信息收益为 $b(0<b<2)$，而双方选择不同信息时收益为 c $(0<c<1)$，当 $0<b<1$ 收益占优均衡时，博弈双方中的一方选择背叛也会获得收益，鼓励个体选择背叛策略；当 $1<b<2$ 风险占优时，一方选择合作也会有所收益，鼓励用户坚持合作策略。从图 5-18 中可以看出，$0<b<1$ 时，随着 c 的增大，合作者比例有所下降；相反，在 $1<b<2$ 区域内，合作比例有所提高。

图 5-19 描述了 p_r，b 和 c 对于 f_c 的影响。当 $b<1.2$ 时，合作者比例一直维持在相当高的水平上，p_r 对 f_c 影响不大。随着 b 的增大，特别是在 $b>1.2$ 的区域内，合作者比例有着明显下降，f_c 随着 p_r 的增大而增大。当 p_r 较小时，合作难以维持而消失。基于声誉社交关系调整 p_r 对合作具有促进作用，声誉作为个体的长期关注点，使得用户谨慎地与社交网络中其他个体进行交互。网络中用户对自身声誉环境越重视，对是否背叛原信息转而采用新信息越谨慎，尤其是在个体对自身声誉重视度较高的情况下，虽然新信息 b 的诱惑不断增大，仍有相当一部分用户选择坚持原信息。而当 p_r 一定时，f_c 随着 b 的增大而减小，背叛诱惑 b 对网

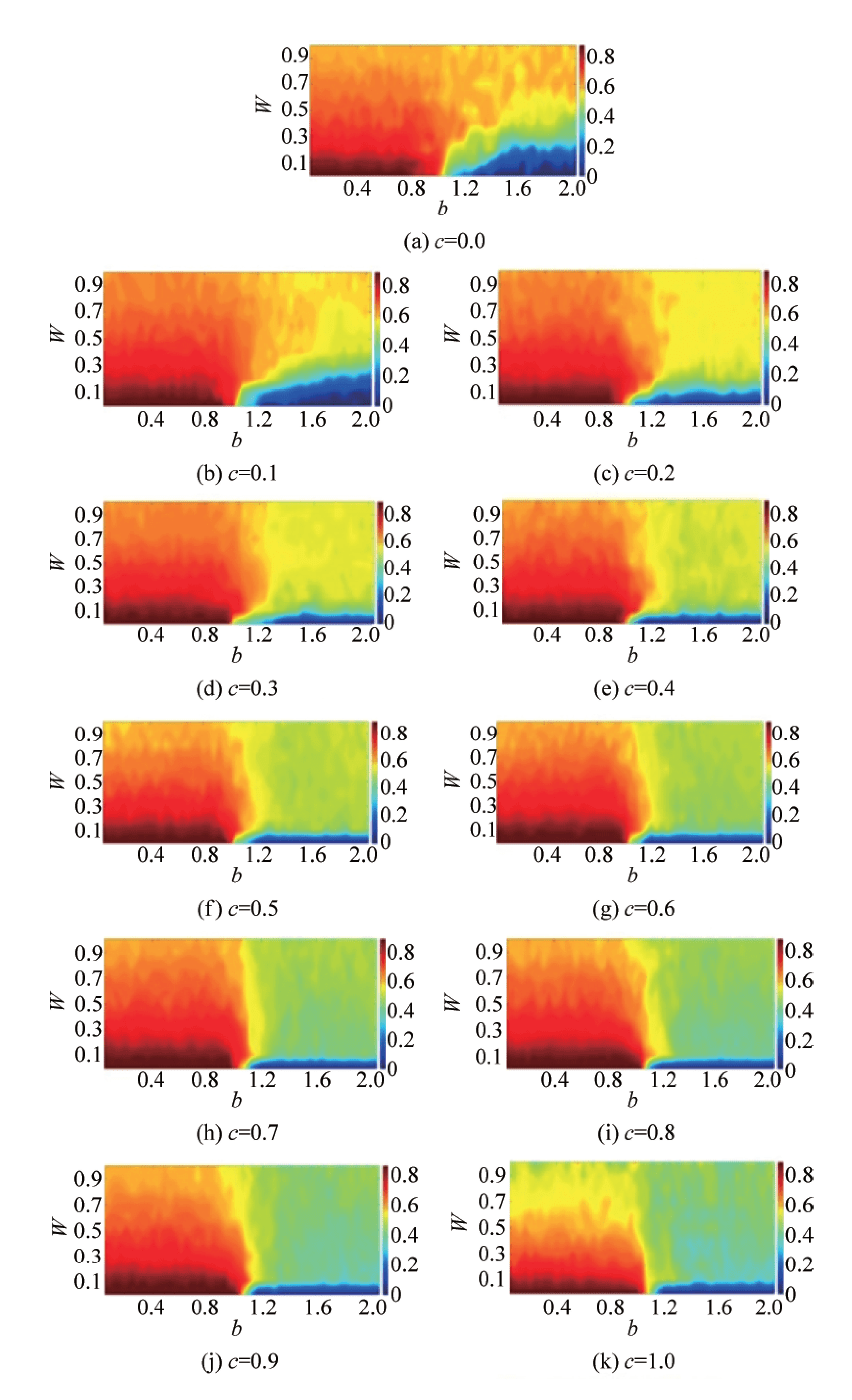


图 5-18　在不同 c 值下，W 和 b 对于合作密度的影响。$p_r=0.8$，$ratio=0.5$

络中合作水平有阻碍作用，随着新信息鼓励背叛的效用 b 的增加，合作变得难以维持，更多的用户放弃原有信息。

与图 5-18 相似，c 的影响作用也体现在当 $0<b<1$ 时，c 对合作有阻碍作用，鼓励用户接收新信息；当 $1<b<2$ 时，c 促进合作的产生，鼓励用户坚持原信息，但最终结果仍然受 p_r 和 W 的限制。

图 5-20 描述了当关注关系更新频率 W 和声誉重视度一定时，$ratio$ 对合作者比例的影

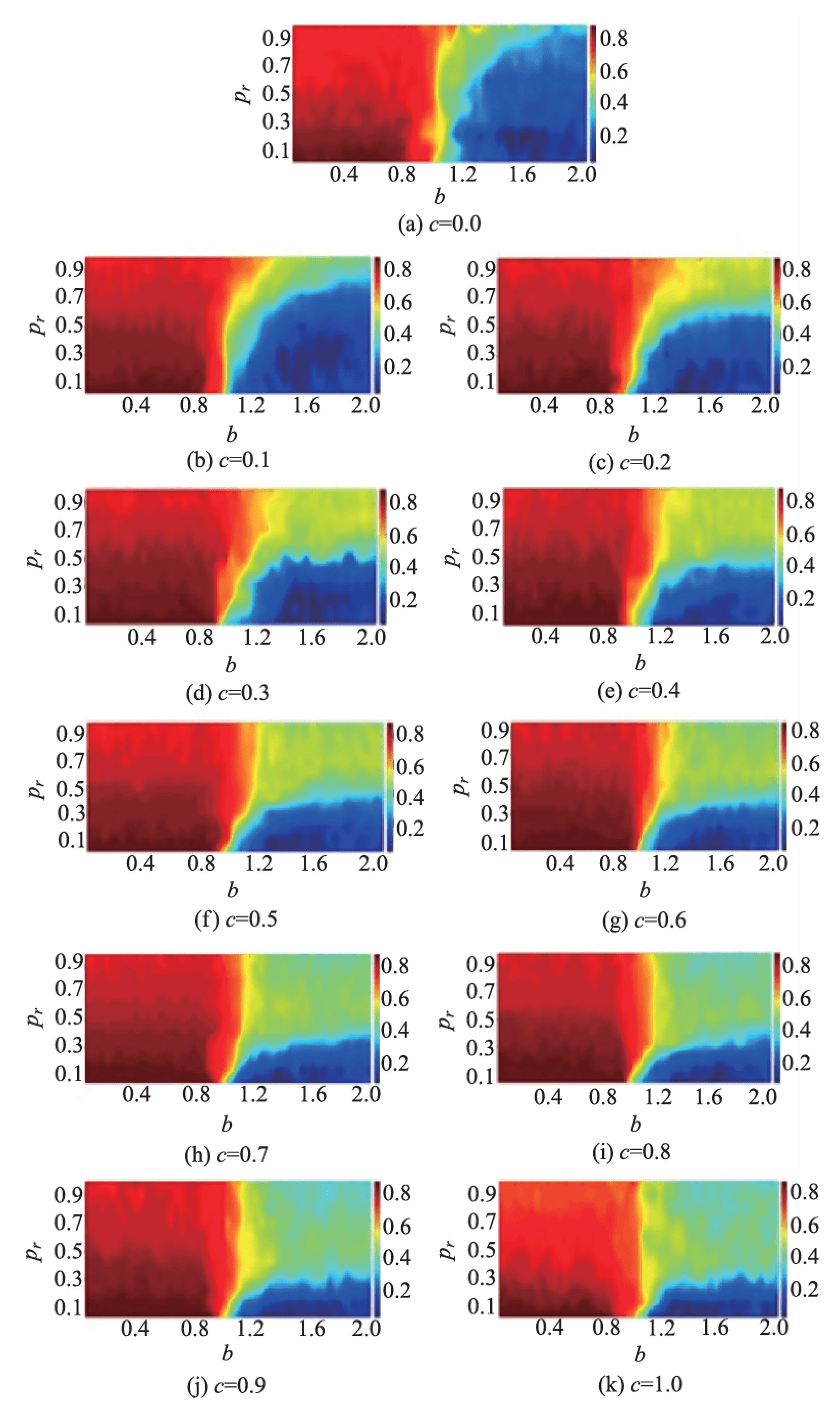


图 5-19　在不同 c 值下，p_r 和 b 对于合作密度的影响。$W=0.2$，$ratio=0.5$

响。由于合作者比例主要受 W，p_r 和 b 的影响，所以取 $c=0.1$ 以方便讨论。当 $ratio$ 接近 0 时，无论 b 如何变化，合作者比例保持在极低水平，网络中几乎没有个体会坚持原信息；当 $ratio$ 接近 1 时，f_c 维持极高水平，无论新信息诱惑有多大也无法改变用户的信息选择。当 $b<1.2$ 时，f_c 随 $ratio$ 的增大而小幅度增大，并维持在较高水平，新信息的诱惑力不足以使用户改变策略选择；当 $b>1.2$ 时，b 和 $ratio$ 的变化对 f_c 的影响较大，f_c 在较低水平上波动，b 值越大，f_c 越小，而 $ratio$ 值仍与 f_c 呈正相关。随着新信息诱惑力的增大，用户开始背叛原有

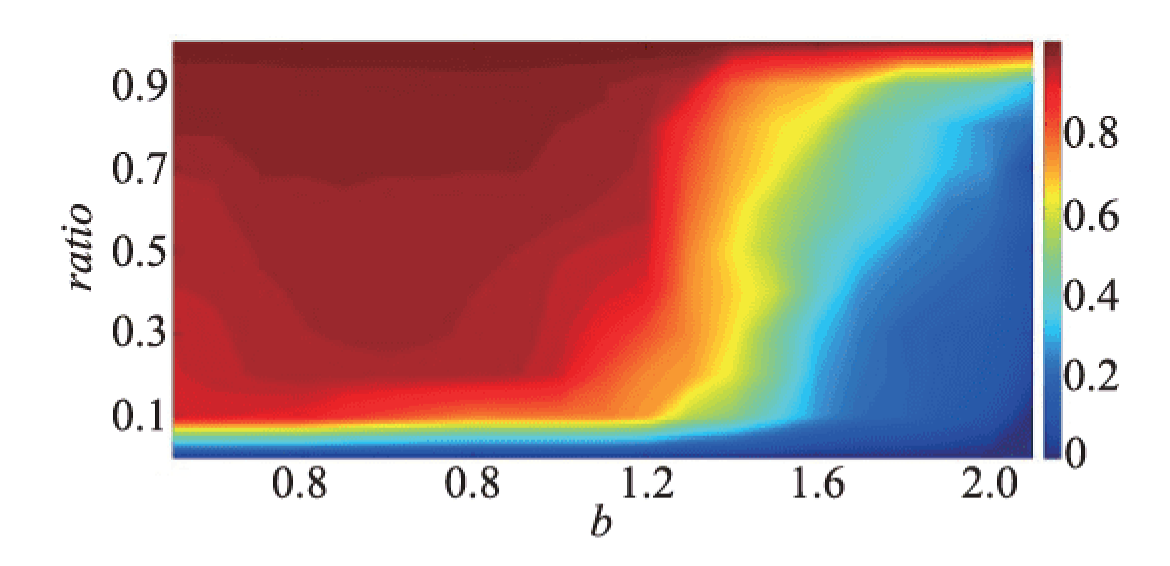


图 5-20 *ratio* 对于合作密度的影响($W=0.2, p_r=0.8, c=0.1$)

信息,而初始群体中用户坚持原信息比例越大,之后转而改变信息选择的用户就会相对减少,这也印证了社交网络中存在的网络效应。

进一步关注不同网络结构对竞争性信息传播的影响。规则网络具有规则图结构,在完全规则网络中,每一个节点都与它的每个邻居节点建立一条连边关系。完全规则网络是稀疏的、连通的,并且具有相对较小的直径、中心节点半径和平均路径长度。这些属性使得规则网络成为研究实际网络的一个很好的选择。在完全随机网络中,每个节点随机选择其他节点,并与之建立连边关系。完全随机网络服从二项式度序列分布的特性,也是网络科学分析中一个重要的研究方向。然而,完全规则网络和完全随机网络都不能再现许多实际网络同时具有的明显聚类和小世界特征。日常生活中接触到的实际网络,大部分既不是完全规则的,也不是完全随机的。在现实生活中,人们经常与他们的邻居和同事交流,但也会联系远在异国他乡的朋友。社交网络上的网页也并非简单地、完全随机地连接在一起。作为从完全规则网络向完全随机网络的过渡,只要在规则网络中引入少许的随机性就可以产生具有小世界特征的网络模型,称为 WS 小世界模型。对于较小的重连概率 p,大多数网络仍保持结构化。因为 k-规则网络具有较高的聚类系数,所以小世界网络也具有较高的聚类系数。但是,即使少量的重连也会大幅度地减少网络直径、半径和平均路径长度,同时保留大多数的 k-规则网络内在的聚类,这称为小世界效应。小世界网络可以清晰地再现多个现实网络中具有的聚类和小世界特性。因此,选取了完全规则网络、完全随机网络和小世界网络这 3 种典型的网络结构来研究在这 3 种网络中竞争性信息传播的差异。

图 5-21 显示了不同网络结构下信息竞争传播的情况。通过调节生成小世界网络函数中随机性参数 p 的值就可以实现从规则网络到随机网络的过渡。$p=0$ 对应于完全规则网络,每个节点都有 k 个邻居节点,完全没有随机跳跃边;$p=1$ 对应于完全随机网络,所有边都是随机重连。由于引入较小的随机性 p 就可以实现小世界网络的聚类和小世界特性,所以模型仿真中参数 p 的取值为 0.03。从纵向对比来看,随着网络结构由完全规则网络到小世界网络,再到完全随机网络,网络的随机性逐步增大,平均合作率演化图中合作程度较低的低温(蓝色)区域逐渐减小。可见网络的随机性在一定程度上促进了个体合作行为的产生。

从上述几幅图中可以直观地发现一个值得关注的现象:因为当 $b<1$ 时为收益占优纳什均衡,而当 $b>1$ 时为风险占优纳什均衡,所以可将 $b=1$ 视为理论上的临界点,即在该点,用户各有 50%的概率选择坚持原信息或者转投新信息。但是从图 5-18~图 5-20 中均能看出,当 b 约等于 1.2 时,相当于新信息产生的收益约为原信息收益的 1.2 倍时用户个体才开始转而偏向选择新信息,也就是说,临界点存在一定的延伸。声誉因素的引入对信息竞争选

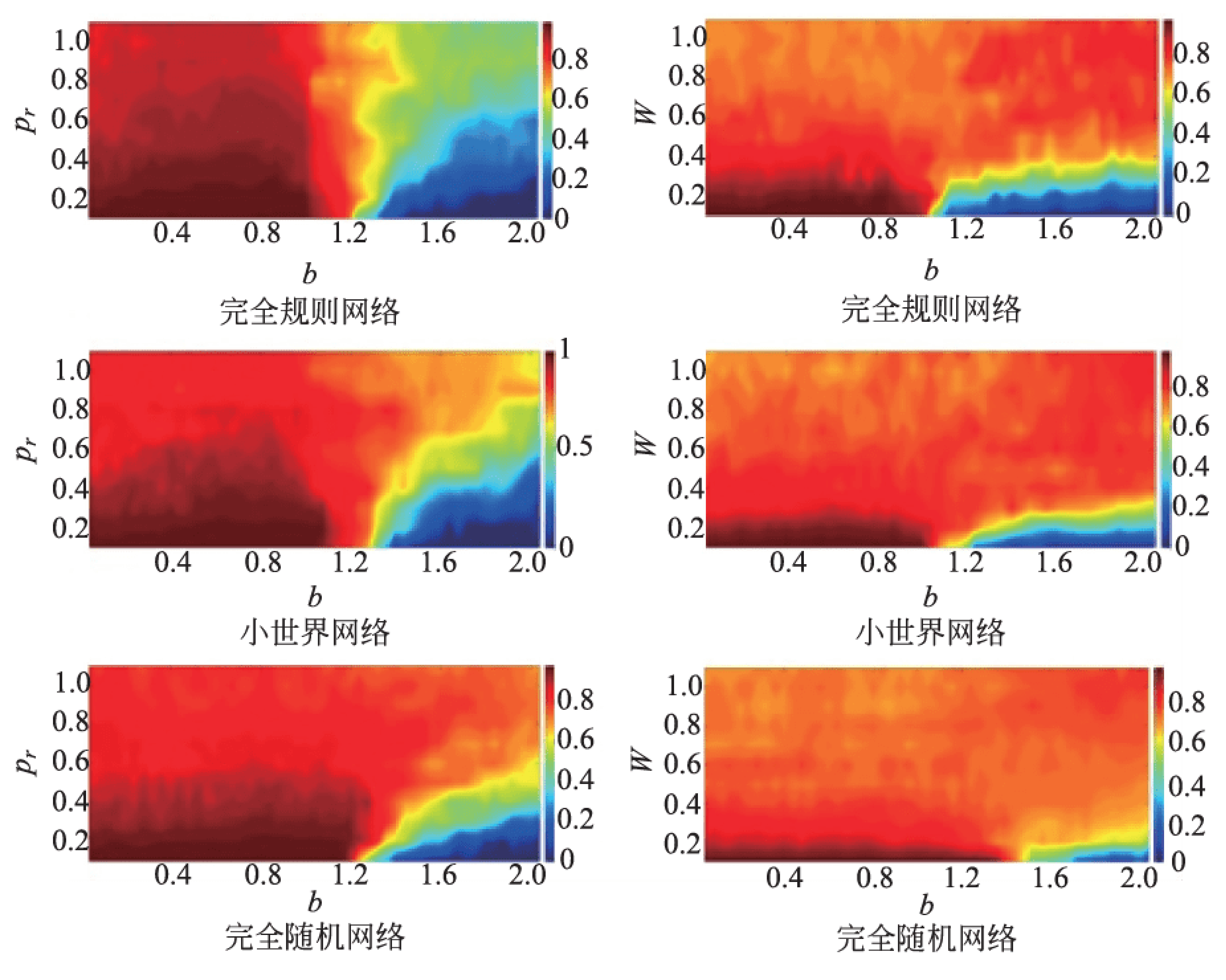


图 5-21　不同网络结构对于合作密度的影响

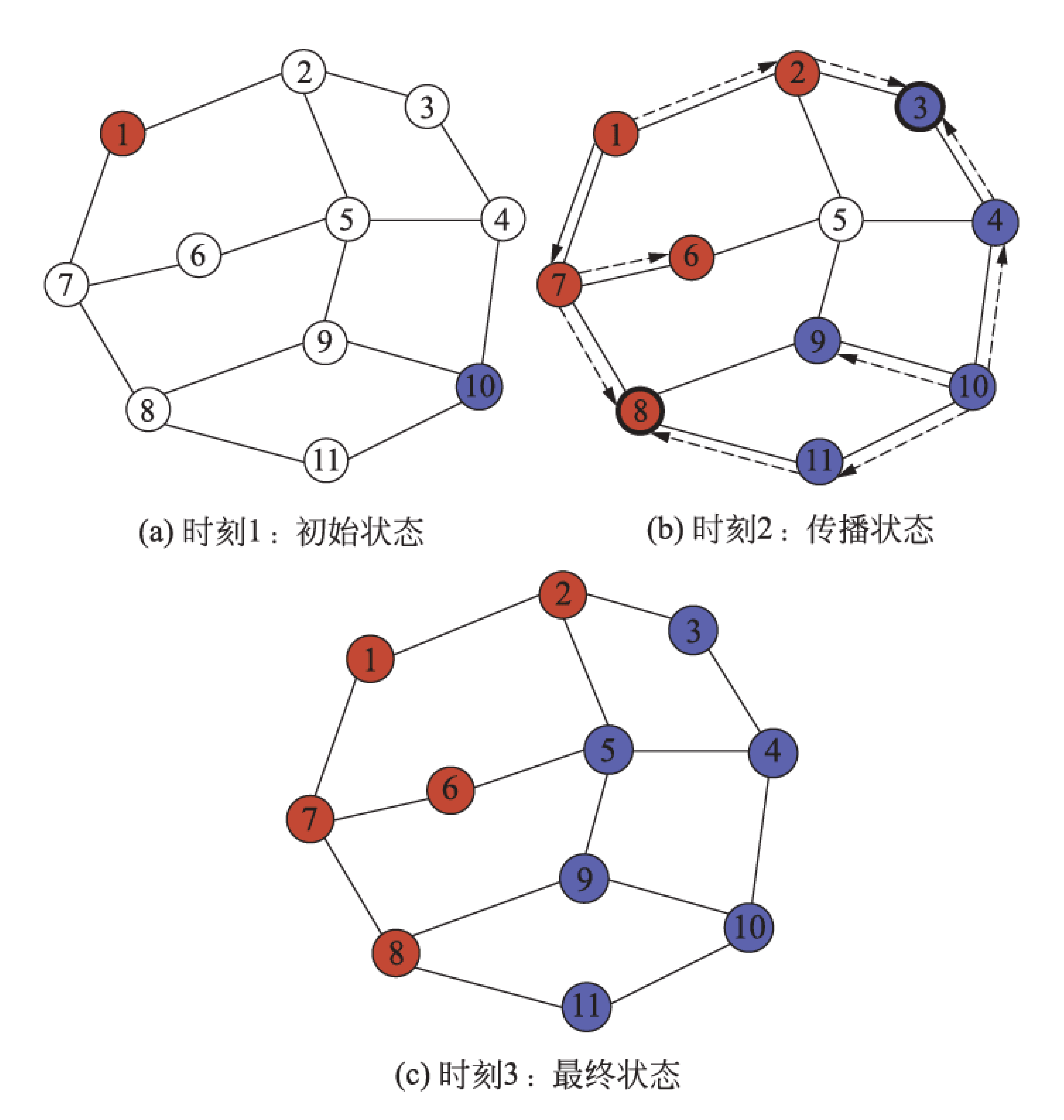
择产生影响，个体不仅关注信息所能带来的收益，还关注改变信息选择会对自身的声誉环境造成怎样的影响。若尽管新信息带来的收益比原信息要大，但仍不能弥补改变自身信息选择对用户所造成的声誉损失，那么用户仍然会坚持原信息。

基于上述分析，小结如下：在基于声誉的信息传播演化网络中，个体对声誉环境重视度、关注关系更新频率以及初始群体合作比例能够促进群体合作发展，用户对改变信息选择更为谨慎，比如只有当新信息收益是原收益的 1.2 倍时，才能弥补用户改变其信息选择对其声誉环境造成的损失。本应用的实际数据分析部分见 11.2 节。

5.2.8.2　竞争信息同步传播演化过程分析

同步信息传播是由社交网络中的用户针对同一时间内，从某一个特定主题所产生的多个相关信息，用户个体与他们的网络邻居用户们(社交网络好友)进行交流和分享相关信息或者意见的过程。而这种信息沟通交流的过程，受到用户个体自身对信息接收程度以及对信息个人偏好的影响，因此可以看作是一个考虑用户个体声誉因素影响、基于信息交互的博弈过程。同步信息传播过程如图 5-22 所示。红色和蓝色分别代表了网络上的两种竞争性信息 A 和信息 B。

图 5-22 中，在线社交网络用一个有向网络来描述，其中每一个圆圈代表网络上的每一个用户个体，圆圈之间的连线表示用户可以根据自身意愿选择建立或者切断与其他用户进行交流的社交关系连边。在传播初始阶段，两个竞争性信息同时加入网络，开始影响用户进行传播，并且尽最大可能来吸引更多的用户关注。网络用户对于是否接纳该信息，不仅出于对信息对自身所产生的吸引力以及带来的收益的考量，而且还受到周围网络用户，即用户的



(a) 时刻1：初始状态　　(b) 时刻2：传播状态

(c) 时刻3：最终状态

图 5-22　同步竞争性信息传播过程

社交好友对于该信息接纳与否的行为影响。图 5-22(b)描述了两个竞争性信息在社交网络上是如何进行传播的。信息价值越高，越能吸引用户的关注。就图示举例说明，随着深色信息 A 和浅色信息 B 在网络上的持续传播，不断地影响着周围用户，虚线箭头表示信息的传播路径和方向。当信息 A 和信息 B 同时传播至用户 3 和用户 8 时，这两位用户必须根据自己的偏好选择其中一类信息。经过权衡比较，用户 3 选择了信息 2，用户 8 选择了信息 A，这便是这一轮信息竞争选择的结果。随着信息竞争演化传播的进行，网络上的信息传播最终达到一个相对稳定的均衡状态，如最终状态图所示。两类竞争性信息在社交网络中以一定比例共存，用户的信息选择以及周围用户如何选择信息的行为，对信息最终的传播结果均会产生一定的影响。

1. 模型设置

竞争性信息同步传播模型也同样基于第 4 章中提出的社交演化博弈(SEG)模型进行搭建。采用无权有向的均匀度随机网络表示社交网络用户之间的社交关系。网络中节点代表实际社交网络中的网络用户，用户个体可以采用合作 $C(s=[1,0]^{\mathrm{T}})$或背叛 $D(s=[0,1]^{\mathrm{T}})$两种策略。个体 i 有两类邻居：入邻居N_i^I与出邻居N_i^O，对应地，有入度 d_i^I和出度 d_i^O。社交演化博弈关系集合表示为 $G=(V,E;I,U,R)$。V 表示个体集合；E 为关系集合；I 为个体间信息行为集合；U 代表效用函数集；R 为声誉函数集。

本节假设社交网络中的个体是非理性且自私的，在与其他用户个体进行协调博弈交互时只关注自身的效用。同时，由于社交网络结构的限制，以及个体缺乏全局信息造成的短视，个体只能获得其所处环境的局部信息，即个体只能获得其邻居节点与邻居节点的邻居节

点的信息。每个博弈个体与其所有连边个体进行协调博弈交互，每个个体的效用是其参与所有博弈所获得的收益之和，即

$$u_i = \sum_{j \in N_i} s_i^{\mathrm{T}} M s_j$$

其中，N_i是个体 i 所有参与博弈的选择合作或者背叛策略的邻居集合；s_i和 s_j 分别为博弈双方个体 i 和个体 j 的策略；M 是所采用协调博弈的 2×2 收益矩阵：$M=\begin{pmatrix}1 & 0\\ 0 & b\end{pmatrix}$。

本文采用双人双策略纯协调博弈(pure coordination game)。在纯协调博弈中，当博弈双方采用相同策略时(同时合作或背叛)，双方收益处于收益矩阵对角线上的纳什均衡状态。当博弈双方采用不同策略时，双方收益均为 0，收益矩阵非对角线均设置为 0，处于非纳什均衡状态。当博弈双方均采用合作策略，即坚持信息 A 时得到的收益为 1；当双方均采用背叛策略，即放弃信息 A 选择信息 B 时得到的收益为 b。其中，$0 \leqslant b \leqslant 2$，当 $0 \leqslant b < 1$ 时，合作得到的收益较大，信息 A 吸引力较大，处于收益占优均衡状态，合作策略占优；当 $1 < b \leqslant 2$ 时，背叛得到的收益较大，信息 B 盈利较大，处于风险占优均衡状态，背叛策略占优。

异步更新机制在第 4 章已经进行了详细描述，这里，针对具体应用场景简要进行描述。在演化和过程的每一步中，每个个体的声誉$R_i(t)$更新为

$$\begin{cases} R_i(t) = \sigma R_i(t-1) + \Delta R_i(t) \\ \Delta R_i(t) = \sin\left(\dfrac{\pi}{2}\dfrac{n_i^C}{d_i}\right) \end{cases}$$

其中，n_i^C表示个体 i 的合作邻居数。sin()函数描述声誉增长速率。

个体策略更新依据：

$$j = \arg\max_{l \in N_i^O} \{u_l > u_i\}$$

模仿获得最大效用的邻居信息行为。

个体社交关系更新以概率p_r为依据：

$$j = \arg \max_{N_i^I \cup_{l \in N_i^O} N_l^O (\{i\} \cup N_i^O)} \{R_l(t) > R_i(t)\}$$

从邻居中选择最大声誉个体建立社交关系，否则，以 $1-p_r$概率随机选择。

同时以概率p_s为依据：

$$k = \arg\min_{l \in N_i^O} \{R_l(t) < R_i(t)\}$$

与邻居中最小声誉个体切断联系。

2. 模型仿真分析

针对社交演化博弈搭建一个具有均匀度的随机网络，网络的平均度 $d=4$，网络中个体总数为 $N=10^3$。演化初始，网络中的个体以相同的概率(50%)选择合作策略或者背叛策略，即选择竞争性信息 A 或信息 B。随着演化的进行，当合作率f_c波动较小时，可以认为协同演化达到相对稳定的状态，否则，继续演化一段时间，演化时间选择 2×10^5 蒙特卡洛(MC)时间步长。每经过10^3MC 时长，进行 20 次独立实验，计算出平均合作率f_c。仿真过程中，模拟噪声 $\beta=0.01$，切断连边概率$p_s=0.01$，声誉衰退率为 1。信息 A 的收益为单位 1，信息 B 的收益以单位 1 为标准用 b 表示，b 的变化范围是 0～2，仿真中 b 的间隔取值为 0.04。

用平均合作率f_c来描述演化中合作行为的信息传播情况，f_c越高，表示信息 A 占据主导，f_c越低，表示信息 B 占据主导。p_r表示对声誉的重视程度，W 反映了个体关注关系更新的频率，关注关系更新频率随 W 的增大而增大，用户个体间的交流更加频繁。p_r和 W 的采样间隔为 0.1，范围为 0～1。

当关系调整频率 W 一定，信息 B 的效用与声誉重视因素p_r同时变化时，博弈演化中平均合作率f_c的变化情况如图 5-23 所示。模型仿真中，W 参数设置为 0.8，表明网络中用户间交流相对较为频繁。当p_r处于较低水平，$b<1.2$ 时，f_c处于较高水平，说明当群体中参与者较为不关注声誉时，在信息 B 的诱惑不够高的情况下，参与者仍然会坚持采用信息 A；当 $b>1.2$ 时，f_c的值非常小，合作行为难以维持，说明当信息 B 的诱惑足够大，参与者开始转而采用信息 B。并且，f_c值随着效用 b 的增大而降低，随着 p_r 的增大而增大，说明信息 B 的诱惑越大，越吸引参与者偏向采用该信息，而参与者对声誉的重视度越高，对是否转而采用另一竞争信息就越谨慎。当基于声誉的关系调整p_r在较高水平时，合作率也维持在一个较高水平，尽管f_c随着 b 的增大而减小，但波动并不剧烈。用户更多考虑声誉因素时并不会更多地改变其社交关系。

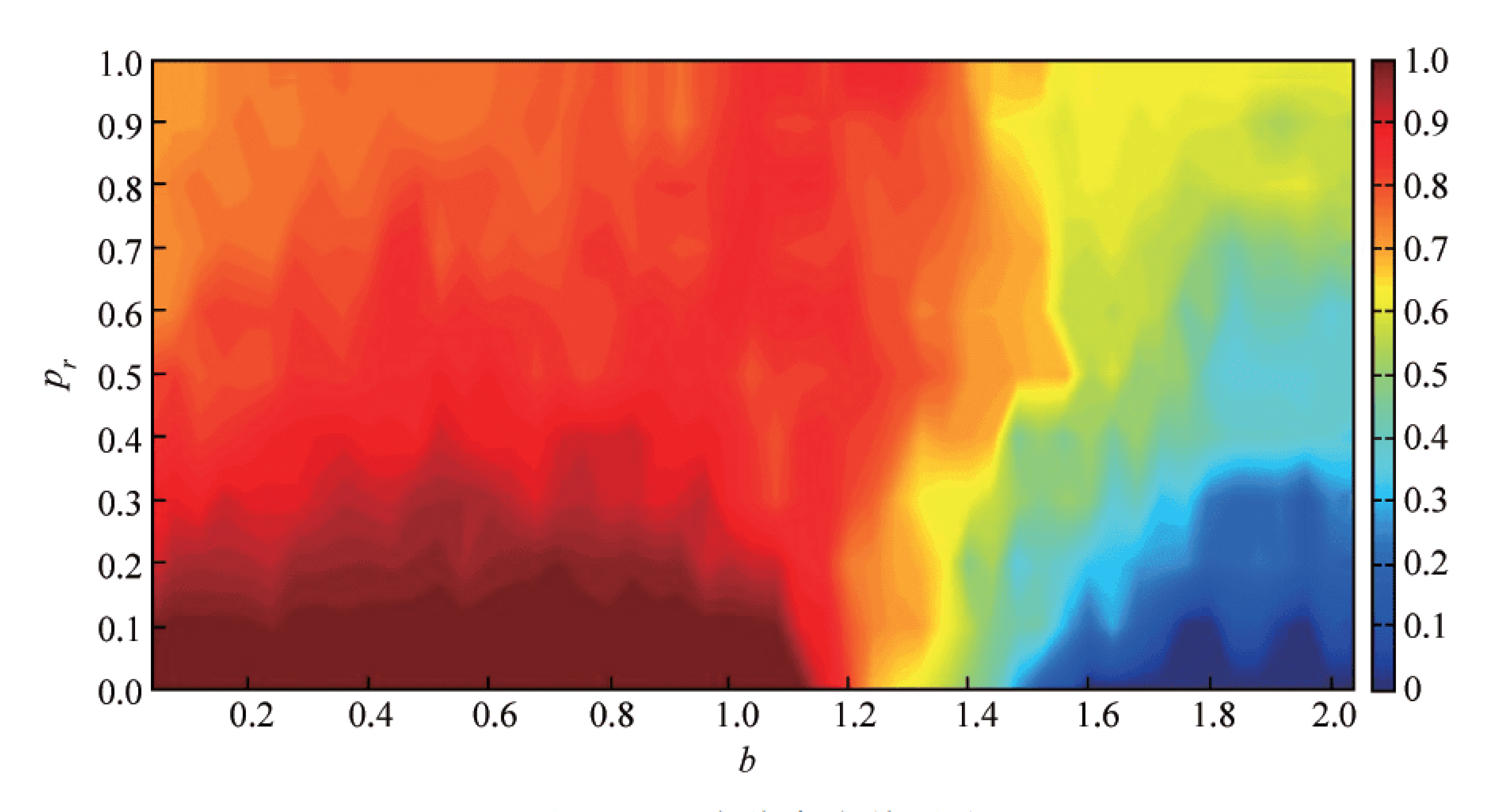


图 5-23　p_r和 b 对于合作密度的影响（$W=0.8$）

结论 1　当用户更重视自身声誉时，对改变其信息选择更为谨慎。

当基于声誉关系调整p_r不变，信息 B 效用与关系调整频率 W 同时变化时，博弈平均合作率如图 5-24 所示。p_r的参数设置为 0.2，说明用户在做信息选择时，声誉因素影响较小。当 W 接近 0，$b<1.2$ 时，合作水平相当高，说明在群体中参与者关系调整频率非常小的情况下，也就是说，与其他参与者交流较少并且信息 B 的诱惑不够高时，参与者仍然会坚持采用信息 A；而当 $b>1.2$ 时，f_c值非常小，说明信息 B 的诱惑足够大，参与者由信息 A 转而采用信息 B。当 W 处于较低水平时，b 值的变化对f_c产生较大影响，f_c随着 b 的增大而减小，当群体中的参与者关系调整频率较低，也就是说，参与者不经常与其他参与者进行交流时，随着信息 B 诱惑的不断增加，参与者逐渐由采用信息 A 转而采用信息 B。可见，当参与者交流较少时，不同信息带来的收益对参与者的信息选择影响较大。而当 W 较高时，信息效用很难再影响合作背叛的行为，用户更多地会坚持合作行为，不轻易改变信息。

结论 2　用户间的频繁交流促进了合作行为的产生，更会坚持已选择的信息。

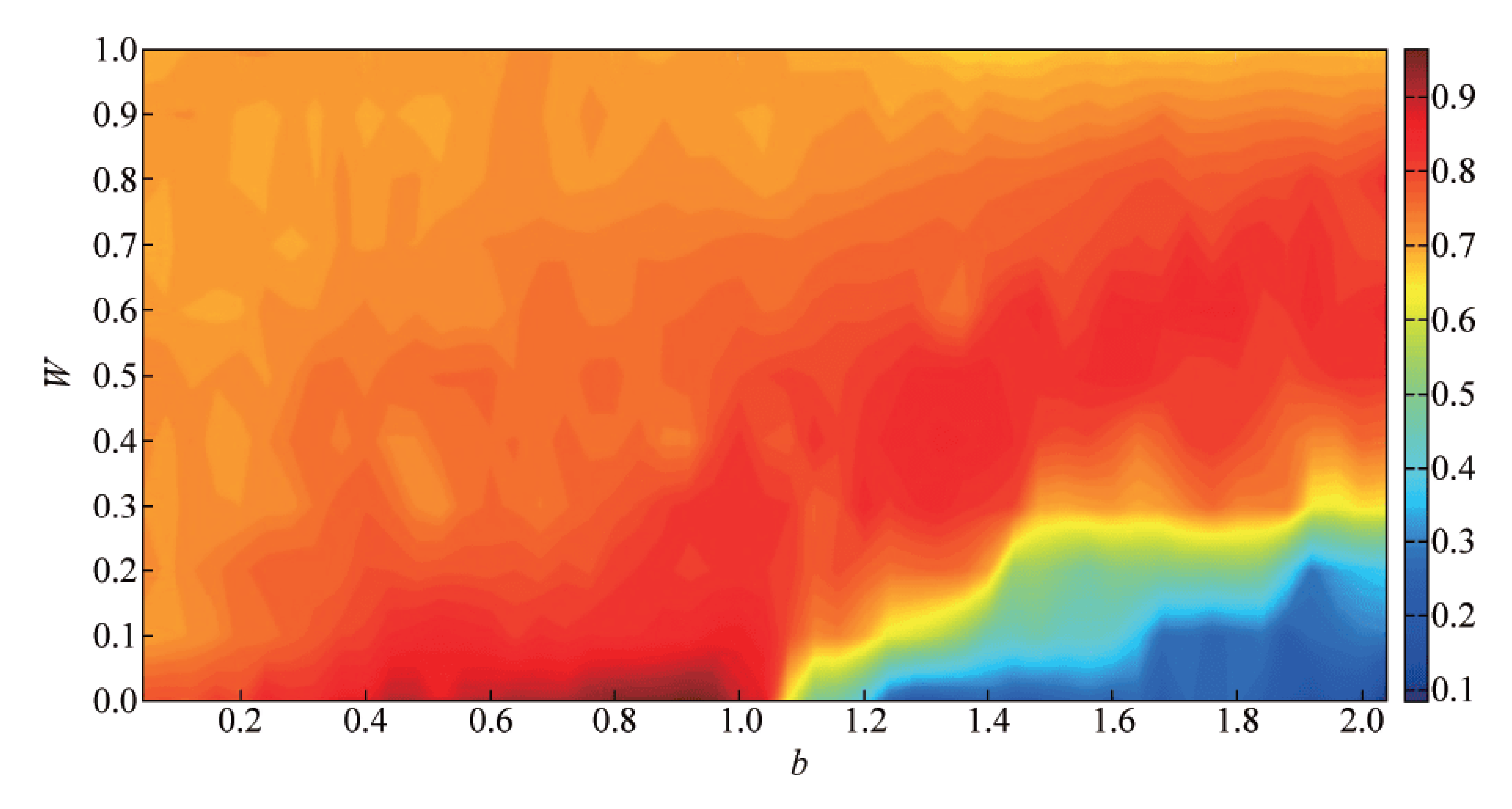


图 5-24　W 和 b 对于合作密度的影响($p_r=0.2$)

从以上仿真图中可以发现一个值得关注的现象。因为 $b<1$ 时为收益占优均衡，$b>1$ 时为风险占优均衡，所以，$b=1$ 可以视为协调博弈一个理论上的临界点，在该点，参与者各有 50%的概率选择信息 A 或信息 B，但从图上可以看出，当 $b=1.2$ 左右时，相当于信息 B 产生的收益约为信息 A 收益的 1.2 倍时参与者才开始由信息 A 转而偏向选择新信息 B，也就是说，临界点有一定的延伸。

从图 5-23 中可以看出，声誉因素对参与者的信息选择产生了一定的影响，参与者在面对信息选择时，不仅仅考虑两种信息所能带来的收益，还考虑不同信息对自身声誉的影响，如果尽管新信息 B 所产生的收益比信息 A 要大，但仍然不能弥补改变自身信息对参与者所造成声誉的损失，那么参与者仍然会坚持信息 A。

从图 5-24 可以看出，参与者关系调整频率对参与者的信息选择产生了一定的影响，也就是说，参与者与他人交流的多少会对信息竞争结果产生影响。当参与者考虑自身声誉时，若参与者与他人的交流较少，则声誉对参与者自身的影响不够大，该参与者也就更可能尝试改变自己的信息，而若参与者经常与其周围的参与者进行交流，那么声誉对他而言较为重要，这时他面临信息选择时就会比较谨慎。

本部分的实际数据分析内容见 11.3 节。

小结

如前所述，社交网络的拓扑结构对信息传播与演化有着重要的影响。另一方面，社交网络上传播的信息存在着关联关系。本章针对拓扑结构对信息传播演化的影响作用以及竞争性信息的传播机理进行了研究。

针对网络拓扑结构对信息传播演化的影响，提出了一个考虑网络连边信息传播能力差异性的一致性模型，重点研究了随机网络和小世界网络哪种网络更适合信息传播。研究结果表明，在不同的异质性强度下，模型所表现的结果完全不同。当异质性很强的时候，小世界网络的信息传播受到很大的抑制，传播的范围变窄，并被随机网络的传播范围超越。而

且,随着异质性的加强,信息在小世界网络上传播的速度变慢,但其生命周期却延长了。当异质性很弱的时候,信息传播在小世界网络上的表现是最好的,并超越了随机网络与规则网络,这与文献[10]和文献[23]的结论一致。然而,异质性对于随机网络并没有影响。

在竞争性信息的传播演化方面,提出了新的竞争信息传播模型,同时对其进行了稳定性分析。分析表明,本章模型所刻画的竞争信息传播过程是稳定的,存在全局平衡点,使得网络信息的传播趋于稳定状态,计算机模拟实验与理论分析得到了相同的结果。针对网络结构对竞争信息传播的影响进行了模拟实验,结果表明,在小世界网络中,后发的抑制信息对待抑制信息的抑制效果更好,在无标度网络中,抑制效果比较差一点。而针对抑制信息的发布时间点对抑制效果影响的模拟实验,发现调控信息的发布时间点的选择对多信息的传播过程存在着重要影响,调控信息发布的时间与已传播信息的传播时间的间隔越小,则对已传播信息的调控效果越好,因此,若想抑制已传播的信息的蔓延,应尽早采取措施,发布强度更大的信息。当过了某个时间节点后,抑制信息的发布时间将对已传播信息的抑制效果不再产生作用。针对网络节点的特征对竞争信息传播的影响作用,模拟实验表明,网络节点的在线频率与节点的度数对传播过程都有影响。抑制信息发布节点的在线频率越高,则最终网络状态控制信息对已传播信息的抑制效果越好。同时,对于发布节点的度数,控制信息的发布节点的度数越大,则在抑制的起始阶段,抑制速度越快,然而网络中某度数节点在网络中出现的数量越多,则最终稳态时抑制效果会更好些。

进一步地,应用社交演化博弈模型研究竞争性信息传播机制的演化过程。对于竞争性信息同步传播机制来说,竞争性信息同时加入网络争夺用户的传播问题;而对于竞争性信息异步传播机制来说,则是竞争性信息先后进入网络来争夺用户扩大影响范围的传播问题。与同步竞争传播不同,异步竞争传播中原有竞争信息对用户的影响较大,先加入竞争传播的信息对用户的影响程度对竞争传播结果起到一定程度的影响。基于社交演化博弈模型,采用一般协调博弈策略来调整更新机制,模拟用户信息选择行为。通过模拟仿真分别分析了两类竞争传播机制及影响因素。

参考文献

[1] Erdos P, Renyi A. On the evolution of random graphs. Publ. Math. Inst. Hungar. Acad. Sci, 1960, 5: 16-61.

[2] Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature, 1998, 393(6684): 440-442.

[3] Newman MEJ, Watts DJ. Renormalization group analysis of the small-world network model. Physics Letters A, 1999, 263(4): 341-346.

[4] Barabási AL, Albert R. Emergence of scaling in random networks. Science, 1999, 286(5439): 509-512.

[5] Smilkov D, Kocarev L. Influence of the network topology on epidemic spreading. Physical Review E, 2012, 85: 016114.

[6] Hébert-Dufresne L, Noël PA, Marceau V, *et al*. Propagation dynamics on networks featuring complex topologies. Physical Review E, 2010, 82: 036115.

[7] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Physical Review Letters, 2001, 86: 3200.

[8] Pastor-Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks. Physical Review E, 2001, 63: 066117.

[9] Volchenkov D, Volchenkova L, Blanchard PH. Epidemic spreading in a variety of scale free networks. Physical Review E, 2002, 66: 046137.

[10] Lv L, Chen DB, Zhou T. The small world yields the most effective information spreading. New Journal of Physics, 2011, 13(12).

[11] Miller JC. Percolation and epidemics in random clustered networks. Physical Review E, 2009, 80: 020901.

[12] Grabowski A, Kosiński RA. Epidemic spreading in a hierarchical social network. Physical Review E, 2004, 70: 031908.

[13] Huang L, Park K, Lai YC. Information propagation on modular networks. Physical Review E, 2006, 73: 035103.

[14] Estrada E, Kalala-Mutombo F, Valverde-Colmeiro A. Epidemic spreading in networks with nonrandom long-range interactions. Physical Review E, 2011, 84: 036110.

[15] Fu F, Liu L, Wang L. Information propagation in hierarchical networks. In: Proc. of the 46th IEEE Conf. on Decision and Control (ICDC'07). 2007. 5329-5333.

[16] Fan P, Li P, Jiang Z, *et al*. Measurement and analysis of topology and information propagation on Sina-microblog. In: Proc. of the 2011 IEEE Int'l Conf. on Intelligence and Security Informatics (ISI'11), 2011. 396-401.

[17] Karsai M, Kivelä M, Pan RK, *et al*. Small but slow world: How network topology and burstiness slow down spreading. Physical Review E, 2011, 83: 025102.

[18] Ou C, Jin X, Wang Y, *et al*. The impact of heterogeneous spreading abilities of network ties on information spreading. In: Proc. of the 2015 IEEE Int'l Conf. on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM). IEEE, 2015.

[19] Weng L, Alessandro F, Alessandro V, *et al*. Competition among memes in a world with limited attention. Scientific Reports, 2012, 2.

[20] Miritello G, Moro E, Lara R. Dynamical strength of social ties in information spreading. Physical Review E, 2011, 83(4): 045102.

[21] Onnela J-P, Saramäki J, Hyvönen J, *et al*. Structure and tie strengths in mobile communication networks. Proc. of the National Academy of Sciences, 2007, 104(18): 7332-7336.

[22] Romero DM, Brendan M, Kleinberg J. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In: Proc. of the 20th Int'l Conf. on World Wide Web. ACM, 2011.

[23] Centola D . The spread of behavior in an online social network experiment. Science, 2010, 329(5996): 1194-1197.

[24] Wei D, Zhou T, Cimini G, *et al*. Effective mechanism for social recommendation of news. Physica A: Statistical Mechanics and its Applications, 2011, 390(11): 2117-2126.

[25] Zhao J, Wu J, Xu K. Weak ties: Subtle role of information diffusion in online social networks. Physical Review E, 2010, 82(1): 016105.

[26] Deepayan C, Wang Y, Wang C, *et al*. Epidemic thresholds in real networks. ACM Trans. on Information and System Security (TISSEC), 2008, 10(4): 1.

[27] Kephart JO, White SR. Measuring and modeling computer virus prevalence. In: Proc. of the 1993 IEEE Computer Society Symp. on Research in Security and Privacy. IEEE, 1993.

[28] Cowan R, Jonard N. Network structure and the diffusion of knowledge. Journal of Economic Dynamics and Control, 2004, 28(8): 1557-1575.

第 6 章 信息行为对网络结构的影响

6.1 概述

近年来，在线社交网络服务得到快速、蓬勃的发展，如新浪微博、Twitter、Facebook 等。在这些社交网络服务中，用户通过信息分享、信息交互的方式来联系彼此。通常，这些信息都是由用户自身产生和发布的，并能够被其好友接收到。当用户的好友接收到这些信息，并且对这些信息感兴趣时，他们就会回应这些信息，例如对接收的消息进行转发、回复、点赞等。一方面，社交网络中的这种信息交互方式反映了用户的社交联系，也满足了他们的社交需求。另一方面，用户在发布和分享消息的时候是需要付出成本的，如时间、费用、精力等。因此，存在某些激励和机制使用户克服成本而持续发消息或者对其他用户的消息做出响应。目前，社交网络中这些现象背后的激励和机制还没有被人们完全了解。

社交网络中个体的典型信息行为主要包括信息分享行为和信息交互行为。在第 4 章中介绍了面向社交网络演化性建模与分析的社交演化博弈模型框架。本章将应用社交演化博弈模型框架来研究和分析在社交网络中典型个体信息行为场景下的社交网络宏观网络结构的演化计算。

6.2 社交网络中典型信息行为分析

6.2.1 信息分享行为

信息分享行为是指个体面向其他多个体的消息发布、转发行为。它是个体间一对多的群交互信息行为，能够使得信息在个体间进行传播，具有方向性。基于信息分享行为的方向性，在研究社交网络中个体信息分享行为的演化性时，本文将社交网络抽象为有向网络。

在常见的社交网络服务中，如新浪微博、Twitter、Facebook 等，用

户可以关注他感兴趣的用户,也可以被对他感兴趣的用户所关注,这种用户间的关注关系具有方向性,即用户发布的消息能够被关注他的所有用户接收。用户的关注者能够对接收到的消息进行转发、评论、点赞等,这些行为不仅能被该用户看到,也能够被该用户的其他关注者所看到,这种由用户及其关注者构成的信息分享的群体交互方式对信息及其传播起到了广播和放大的作用。在每一轮的群体信息交互中,用户总是期望其获得收益的最大化,从而具有更高的效用。因此,用户会学习和模仿其成功好友的行为和策略。同时,在社交网络中具有良好声誉的用户总是会受到其他用户的欢迎,进而成为拥有大量粉丝和高影响力的关键节点用户,例如新浪微博中的草根大 V 大多数就是从普通账户开始通过培养良好的声誉获得大量的粉丝和较高的影响力的。通过以上分析可以看到,在社交网络中,由于用户对于行为的效用和所处的关系环境的关注,使得不仅是用户的行为策略,用户之间的关系也会随着时间发生改变。因此,社交网络的演化包括两部分:用户行为策略的演化和用户关系的演化。它们之间是相互影响、相互作用、共同演化的。这种用户行为与用户关系相互影响、共同演化可以引入网络上的演化博弈论中的共演化机制进行研究。

6.2.2 信息交互行为

信息交互行为是指一个个体与另一个个体之间的基于消息的回复、评论等行为。它是个体间一对一的两两交互信息行为,是个体间基于信息而产生的双向交互行为。基于信息交互行为的双向性,在研究社交网络个体信息交互行为的演化性时,本文将社交网络抽象为无向网络。

作为新兴的交流服务平台,社交网络服务,如 Twitter、Facebook、新浪微博等,吸引了全世界数以亿计的用户。每一天,使用这些服务的用户都产生了大量的信息交互行为,比如与好友讨论彼此感兴趣的话题、更新状态信息、与兴趣相似的用户建立联系、互赞以维持好友关系等。因此,社交网络的结构可以被认为是由相对稳定的关系(如 Twitter 中的关注关系)和临时关系(如用户之间的消息)构成,这种结构能够反映出用户关系的远近。由于用户的内在和外在因素,社交网络用户间的社交网络会随着用户的交互而不断地产生和消失,这就使得社交网络的结构具有动态变化的特征。因此,为了能够跟踪社交网络的发展状况,研究和理解社交网络动态演化的内在本质具有重要的意义。

6.3 网络结构更新方式分析

6.3.1 基于随机选择的网络结构更新方式

在社交网络中,最简单的用户关系更新方式是依概率随机进行用户关系更新。在这种用户关系更新机制下,通常依给定概率随机选择一条给定类型的关系,然后判断该关系是否需要更新,如果进行更新则断开该关系,该关系的两个个体中的一个个体依概率随机选择新的个体建立关系;或者从网络中随机选择一个个体,依给定概率随机选择一个个体建立新的关系。Zimmermann 等人提出了一种只影响“背叛者-背叛者”用户关系的用户关系更新机制,他们认为,这类关系中的用户应该寻找新的伙伴以获得更好的效用。他们的研究显示,在初始网络为给定平均度的随机网络和最富者跟随策略模仿规则下,即使以很小的概率 p

来寻找新的伙伴用户，系统仍然能够促进合作。同时，对于网络结构来说，该用户关系更新机制能够形成层次交互结构，如果新伙伴用户被限制在二度邻居中，则可能使网络生成小世界网络属性。但是，这种随机断开“背叛者-背叛者”关系的方式并不能使网络的度分布形成无标度分布。后续的研究[2,3]详细说明了高收益能够带来的社交网络中合作者的自发涌现，并且被称为“领导者(leader)”的个体在全局合作维持中扮演着重要的角色。因此，这些研究工作初步揭示了社交网络通常所拥有的无标度特性是非常适合合作演化的。

6.3.2　基于声誉的网络结构更新方式

在上一节中可以看到，社交网络中“领导者”这一角色在社交网络的演化中扮演着重要角色。在社交网络中，高声誉个体就是“领导者”中的一种个体。通常，一个个体的声誉是指其他个体对该个体的综合评价，如该个体的信息行为的合作程度。基于声誉的用户关系更新方式是指个体在进行关系更新时会选择声誉度高的个体作为新的伙伴，同时舍弃声誉度低的伙伴。Fu 等人[4]研究了声誉在社交网络演化中的作用和影响力。在他们的模型中，社交网络中的个体只能获得二度邻居的信息，因此，当个体进行关系更新时，他依概率 p_r 选择二度邻居中声誉最高的个体作为新的伙伴，或者依概率 $1-p_r$ 随机地从社交网络中选择一个个体作为新的伙伴。同时，断开与其邻居中声誉最低的个体的关系。在社交网络演化的过程中，每当一个个体进行一次合作行为时，其声誉值就会加 1，否则，不变。这样，个体的声誉度就反映了该个体在演化历史中的合作程度。在该模型下，社交网络从给定平均度的随机网络演化生成高度异质网络，该生成网络的度分布呈现幂律分布。高度数个体通常是高声誉的合作个体。这显示了声誉和度之间的正相关性，提供了一种正反馈机制，该机制能够有效地促进社交网络的合作水平。

6.3.3　基于影响力的网络结构更新方式

社交网络中的“领导者”除了高声誉个体外，还有一种个体——影响力个体。通常，一个个体的影响力是指该个体对于其他个体施加影响的能力，或者是个体能影响到其他个体的程度，如社交网络中的高度数个体、新浪微博中的认证大 V 等。基于影响力的用户关系更新方式是指个体在更新关系时选择有影响力的个体作为伙伴建立关系。这种更新方式是一种富者跟随(richest following)机制，典型的模型是择优连接模型(preferential attachment model)[5]。在该模型中，①增长：初始时，网络有 m_0 个节点。此后，在每一个演化时间步内都有一个新的节点到来，并和网络中的 m 个不同节点建立连接($m<m_0$)；②择优连接：网络中已存在的节点 i 被新节点选中建立连接的概率正比于该节点的度 d_i。该模型的一个显著特点是，它所生成的网络是一个不断增长和演化的网络，并能够刻画社交网络中的无标度特性，从而得到了广泛的应用和研究。

6.3.4　基于关系强度的网络结构更新方式

社交网络中用户间关系强度是指用户之间连接关系的紧密程度。通常，基于关系强度的用户关系更新方式会根据关系的强度来选择是否断开已有的关系，并建立新的关系。反映用户关系程度的一个指标是用户关系的生命时长。Pacheco 等人[6,7]研究了关系生命时长对于社交网络演化的影响。在该模型中，使用合作策略 C 的个体和使用背叛策略 D 的个

体分别被给定形成关系的概率 α_C 和 α_D。这样，关系 xy 形成的比率是 $\alpha_C\alpha_D, x, y \in [C, D]$。同样地，根据每条关系所连接的个体使用的策略，每条关系都被设定了生命时长 $\tau_{xy} = \frac{1}{\gamma_{xy}}$，$\gamma_{xy}$ 是相应关系的死亡比率。根据以上设置，通过使用平均场方程，作者们提出了一种被称为“活跃连接(active linking)”的动态用户关系更新机制。研究结果显示，当活跃连接机制的时间长度远小于策略更新的时间长度时，该动态用户关系更新机制会促使博弈的收益矩阵发生变形，例如使囚徒困境博弈转化为协调博弈，而雪堆博弈则转化为融洽博弈。以上这两种情况都能够促进社交网络中的合作。该模型演化生成网络的度分布呈现快速衰减的特点[8]。

6.4 信息分享行为对网络结构的影响

本文使用无权有向网络表示在线社交网络服务中具有有向性的用户间的关注关系，网络中的节点表示社交网络中的用户。在有向网络中，节点 i 有两类关注关系，即 i 的关注者和被关注者。如果从节点 j 到节点 i 存在关注关系 e_{ij}，则 j 是 i 的关注者。反过来，i 是 j 的被关注者。N_i^I 表示 i 的关注者集，N_i^O 表示 i 的被关注者集。d_i^I，d_i^O 分别是 i 的入度和出度，分别表示 i 的关注者数和被关注者数。d_i 是 i 的度，它是 i 的入度和出度之和。当 i 发布或转发消息时，只有它的关注者才能够接收到。此处假设，当 i 的关注者接收到他的消息时，他们能够意识到消息的内容。

以下将对社交演化博弈模型的博弈规则和共演化更新规则进行详细的介绍，同时，短期效用与长期声誉在社交演化博弈中的作用及其之间的相互影响也将一并介绍。

6.4.1 信息分享行为模型

在线社交网络服务中，基于用户信息分享行为的用户交互具有如下特点。

(1) 用户发布的信息被他和他的关注用户所组成的群体分享，而群体成员对于消息的响应，如转发、评论、点赞等，也会被群体成员分享；

(2) 用户在发布消息、评论时需要付出一定的成本，如时间、精力、费用等；另一方面，用户的这些行为也会带来收益，如获得信息、促进社交关系等；

(3) 存在搭便车现象(free rider)，即用户对信息的只获取不付出行为。

理论上，用户搭便车的行为会导致“公共品悲剧”问题，这会引起用户活跃度的下降，直至没有用户活动。这种由用户及其关注者组成的群体所进行的群体交互行为可以使用公共品博弈进行建模。

在公共品博弈中，个体有两个可选的行为策略，C 和 D。其中，C 表示合作，是指该行为策略能够促进和维持网络中节点的活跃度，信息分享活动持续不断地发生；而 D 则表示不合作，采用该行为策略的节点除了消费信息外没有任何贡献。每个节点都能够主持公共品博弈，或参与其被关注者主持的公共品博弈。当节点 i 主持一个公共品博弈时，所有参与博弈的节点的收益如下：

$$\begin{cases} p_i^D = \eta n_i^C b \\ p_i^C = \eta (n_i^C - 1) b - c \end{cases}$$

其中，i 主持该公共品博弈；n_i^C 表示参与该公共品博弈中采用 C 策略的节点数；η 是该群体的放大效应因子；b 是合作行为提供的收益；c 则是合作行为的付出成本。在社交网络信息分享的语境下，节点提供的收益 b 被群体成员无损地分享，即每个博弈的参与者都能获得每个合作者提供的收益 b。

节点 i 不仅能够主持公共品博弈，也能够参与他的被关注者所主持的公共品博弈，因此，他的效用是其参与的所有公共品博弈的收益之和，即他主持的与参与的博弈所得的收益之和，表示如下：

$$u_i = p_i^{s_i} + \sum_{j \in N_i^O} p_j^{s_j}$$

其中，s_i 是 i 使用的策略。

6.4.2 更新机制

社交演化博弈的演化规则采用异步更新方式，即在每一个演化周期 t，网络中随机的一个节点进入更新状态。共演化更新规则由两部分组成：节点的行为策略更新规则和关注关系更新规则。设在演化周期 t，节点 i 进入更新状态，则他要么以概率 $\frac{1}{1+W}$ 进行行为策略更新，要么以概率 $\frac{W}{1+W}$ 进行关注关系更新。W 是行为策略更新时长 τ_e 与关注关系更新时长 τ_p 之比，即 $W=\frac{\tau_e}{\tau_p}$，该参数反映了节点进行关注关系更新的频率。明显地，关注关系更新的频率随着 W 的增大而增大。

在每一个演化周期 t 内，每个节点的声誉都会进行更新。节点声誉的更新程度依赖于节点自身的行为策略和所处的环境，更新方式如下所示：

$$R_i(t) = \sigma R_i(t-1) + \Delta R_i(t)$$

$$\Delta R_i(t) = \sin\left(\frac{\pi}{2} \frac{(n_i^I(C) + n_i^O(C))}{d_i}\right)$$

其中，$n_i^I(C)$ 表示在 i 主持博弈时他的关注者中采用 C 策略的个数，$n_i^O(C)$ 表示在 i 参与的博弈中其被关注者中采用 C 策略的个数。

(1) 行为策略更新规则：在一个演化周期内，当节点 i 要改进他在群体交互中的短期效用时，他将进行行为策略更新。在 i 进行行为策略更新时，他通过模仿其关注者中获得最大短期效用的关注者所使用的行为策略来改进自己的短期效用，即寻找满足的节点 j，并模仿 j 的行为策略：

$$j = \underset{l \in N_i^O}{\operatorname{argmax}}\{u_l > u_i\}$$

行为策略更新的方式遵循 Fermi 规则：

$$\omega(s_i \leftarrow s_j) = \frac{1}{1 + \exp[\beta(u_i - u_j)]}$$

其中，β 表示节点的非理性选择程度，即节点模仿其他节点行为策略的意愿程度。

(2) 关注关系更新规则：在一个演化周期内，当节点 i 要改进他的长期声誉时，他将进行关注关系更新。在 i 进行行为关注关系更新时，节点通过与高声誉节点建立关注关系并断开已有的与低声誉节点的关注关系来改善自身所处的环境，即与易合作的节点建立联系，

以此来获得自身长期声誉的改进。由于网络结构和节点获取信息能力的局限性，节点只能感知到他所处的局部环境的信息。具体地，当 i 进行关注关系更新时，i 进行两种更新操作：建立新的关注关系和断开已有的关注关系，即 i 以概率p_r按照最大声誉原则，从他的关注者和被关注者的被关注者中寻找节点 k：

$$k = \arg \max_{\mathcal{N}_i^I \cup_{l \in \mathcal{N}_i^O} \setminus (\{i\} \cup \mathcal{N}_i^O)} \{R_l(t) > R_i(t)\}$$

并与之建立关系e_{ik}；以概率 $1-p_r$随机地从 i 的关注者和被关注者的被关注者中选择 k'建立关系$e_{jk'}$。同时，i 以概率p_r取消与其被关注者中声誉最小节点 j 的关系：

$$j = \underset{l \in \mathcal{N}_i^O}{\operatorname{argmin}} \{R_l(t) < R_i(t)\}$$

在社交网络的演化过程中，网络中的节点通过不断的行为策略更新和关注关系更新来改进自身的短期效用和长期声誉。在此过程中，节点的短期效用和长期声誉相互影响，促使了节点行为策略与网络结构的共同演化。

6.4.3　实验结果分析

6.4.3.1　模型仿真结果

本节对上节的基于信息分享行为的社交演化博弈模型进行了数值仿真分析，以研究社交网络上节点的信息分享行为和网络结构的演化。初始时，网络结构为随机有向网络，节点的度为 $d=4$，总的节点个数为10^3。使用合作者比例f_c作为指标来评价节点的信息分享行为的演化。当演化过程满足以下 3 个条件之一时，认为网络的演化进入相对稳态：节点全部合作、节点全部不合作，或者在10^4个演化周期内f_c的波动处于非常小的区间。如果社交演化博弈无法达到相对稳态，则它将演化相对较长的周期数，如 2×10^5。在社交演化博弈进入相对稳态或演化了 2×10^5周期后，再演化10^3个周期，最终得到f_c是这10^3个演化周期的合作者比例的平均值。数值仿真结果是 48 次独立运行的仿真结果的平均值。

影响社交演化博弈演化的重要参数有 3 个，分别是 W、p_r和 η。W 是节点行为策略更新与关注关系更新的时长比例，反映了节点关注关系更新的频率，即关注关系更新的频率随 W 的增大而增大。当 $W=0$ 时，没有任何关注关系更新，只有节点的行为策略更新，社交演化博弈退化为静态网络上的节点行为策略演化。随着 W 的增大，关注关系更新的频率也增大。W 的取值范围为 0～1，步长为 0.1。通常在在线社交复杂网络中，大多数用户调整关注关系的频率并不高。因此，W 的最大值设为 1，即节点进行关注关系更新和行为策略更新的比例相等。p_r反映了节点获得更好的声誉环境的渴望程度。p_r的取值范围为 0～1，步长为 0.1。当$p_r=0$ 时，节点完全不关心声誉，只是随机地选择新节点进行关注；而$p_r=1$ 则表示节点在选择新节点进行关注时完全遵循最大声誉原则。η 表示了群体行为的放大效应，它直接作用在节点的效用上，能够影响节点的行为策略更新。η 的取值范围从 0.4～3，步长为 0.2。其他参数设置如下：节点的非理性选择参数 $\beta=0.01$，断开已有关系的概率 $p_s=0.01$，声誉的遗忘因子 $\sigma=1$，它表示节点能够完全获知其伙伴的历史行为信息。通常，$b=c=1$。

图 6-1 显示了在不同的 W 下，f_c作为(p_r,η)函数时受这两个参数的影响情况。从图中可以看到，随着 W 的增加，f_c波动范围的上界减小，而下界增大。f_c波动范围的上界随 W

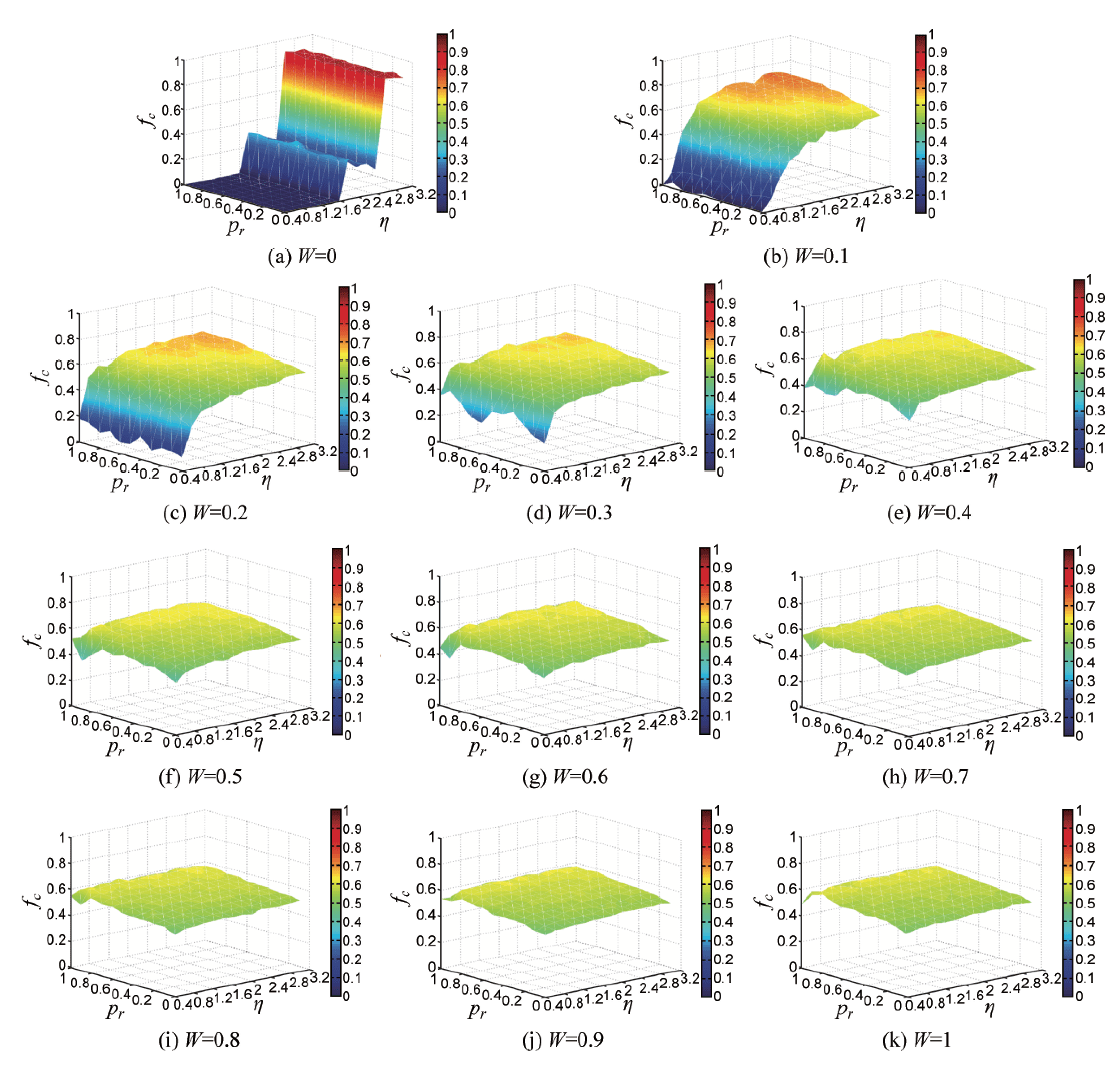


图 6-1　行为策略更新与关注关系更新的时长比例 W 对 f_c 的影响。$N=1000$，$\langle k\rangle=4$，$\sigma=1$，$\beta=0.1$，$p_s=0.01$

的增大而变小的现象说明了关注关系更新频率的增大对于合作者比例的上限具有抑制作用，类似的现象在已有的工作中也有发现。然而，关注关系频率的增大却可以提高合作者比例的下限。换言之，虽然网络中的合作者比例的上限被抑制了，但是合作者却更容易生存。关注关系更新频率对于 f_c 波动范围影响的两面性说明，在社交网络中盲目地鼓励用户频繁地调整关注关系可能并不是一个好的手段。这是因为，虽然它会促使用户活跃度的下限提高，使得用户更容易发帖、评论等，但是也会导致用户活跃度的上限被抑制，更容易碰到天花板。因此，需要恰当、谨慎地鼓励用户进行关注关系的调整。

图 6-2 则着重描述参数 p_r 和 η 对于 f_c 的影响情况。首先关注的是声誉在关注关系更新中的影响，即 p_r 的影响。从图中可以看到，当 W 和 η 确定时，合作者比例 f_c 随着 p_r 的增大而增大，即基于声誉的关注关系更新规则能够促进网络中的合作水平。同样地，η 的增大对合作也有促进作用。若群体的放大效应越能促使节点获得更多的效用，合作者就越容易产生和生存。图 6-1(a)和图 6-2(a)显示的是 $W=0$ 时演化发生在静态网络的情况。在此情况

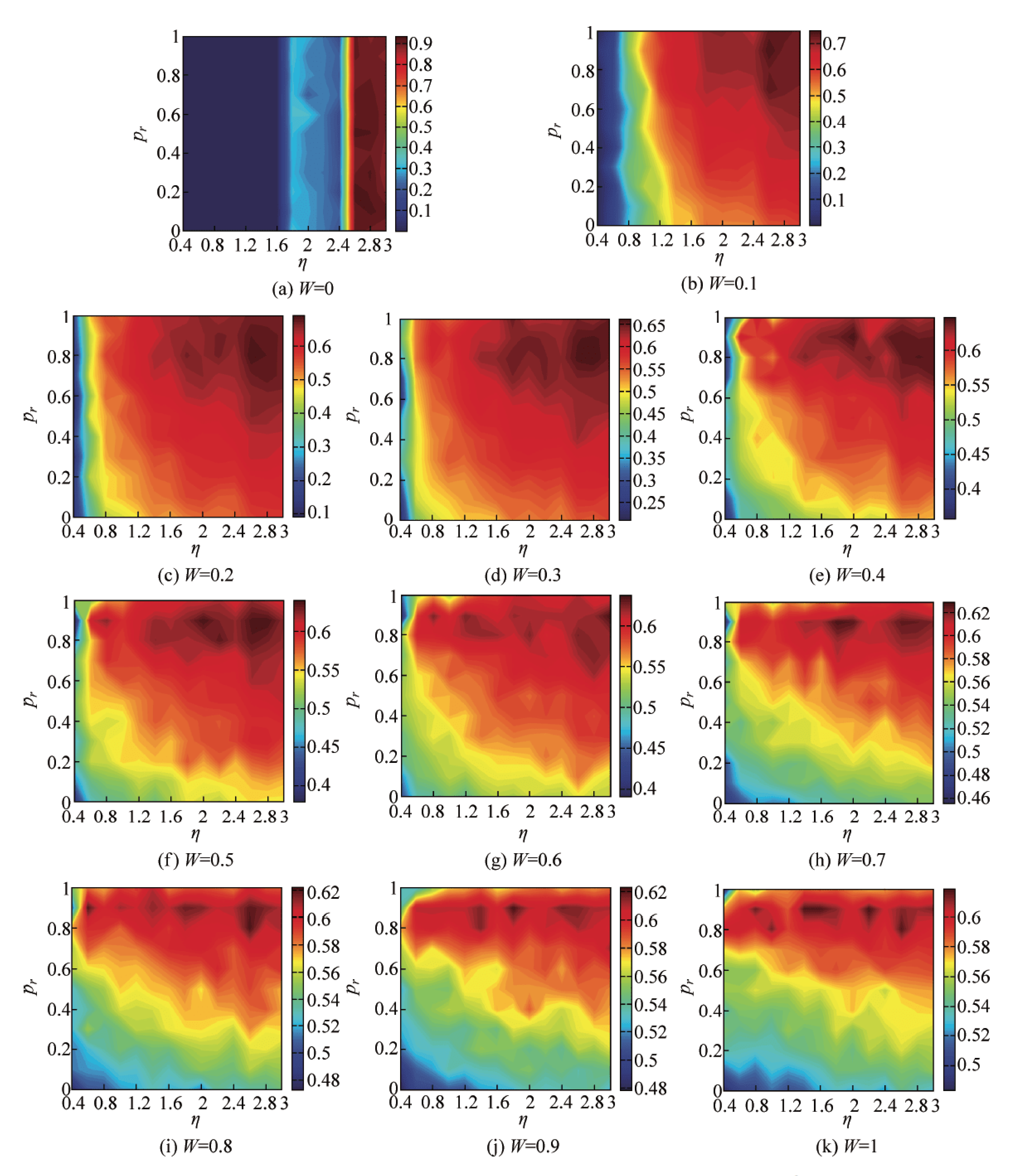


图 6-2 在不同 W 下，p_r和 η 对于合作者比例的影响。$N=1000$，$\langle k\rangle=4$，$\sigma=1$，$\beta=0.1$，$p_s=0.01$

下，合作者比例f_c只受 η 的影响，W 和p_r不起任何作用。

总而言之，随着 W 的增加，合作者越容易产生和生存，但是过大的 W 会使得网络中的合作水平难以达到高水平。增加p_r和 η 可以促进网络中的合作，但是最终的结果都受 W 的限制。

6.4.3.2 真实数据分析

本节应用社交演化博弈模型，利用新浪微博的真实数据的分析结果，对新浪微博中用户

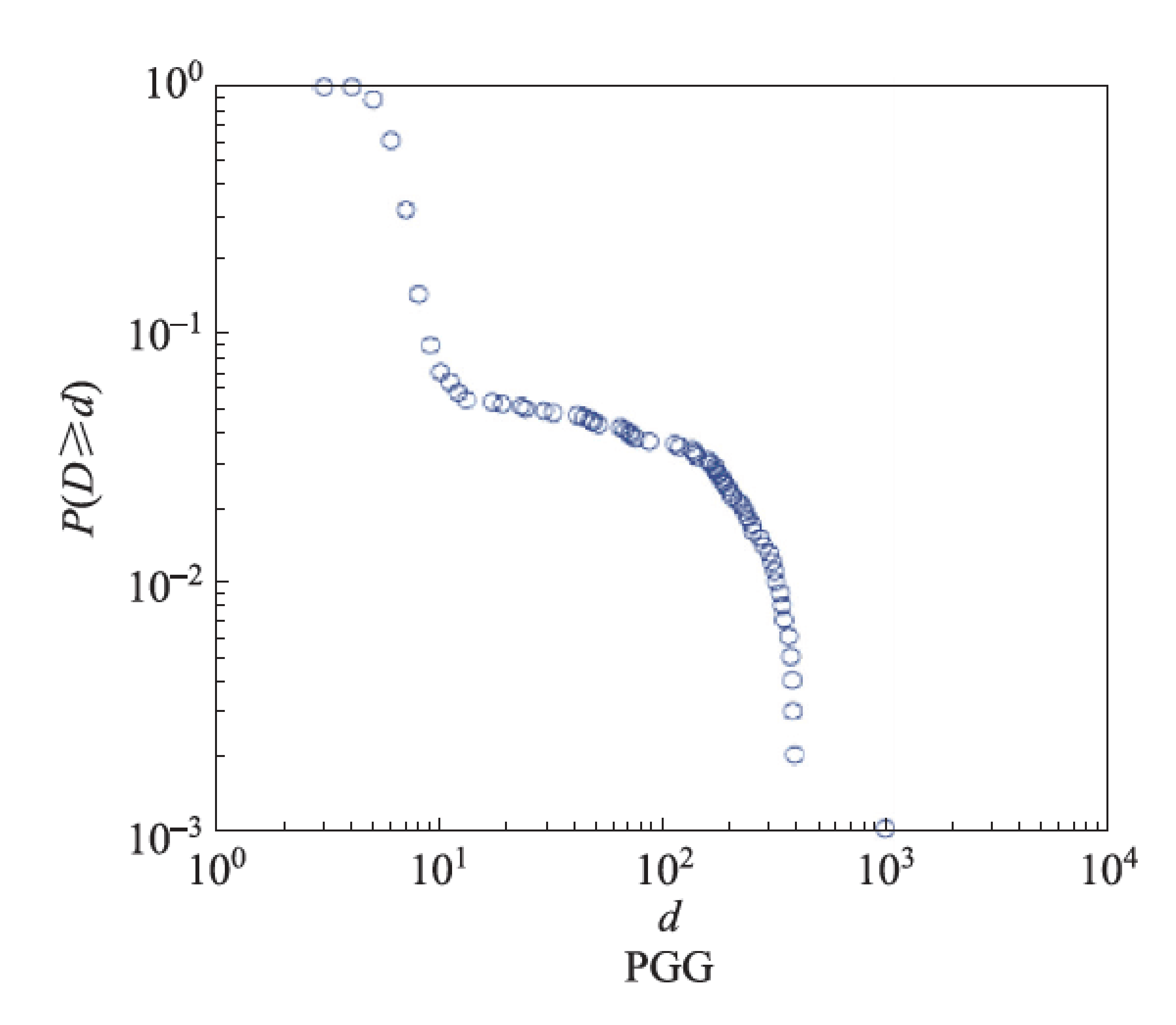


图 6-3　演化生成网络累积度分布。参数设置：$p_r=0.2$，$N=10^3$。采样时间：$t=2\times10^5$

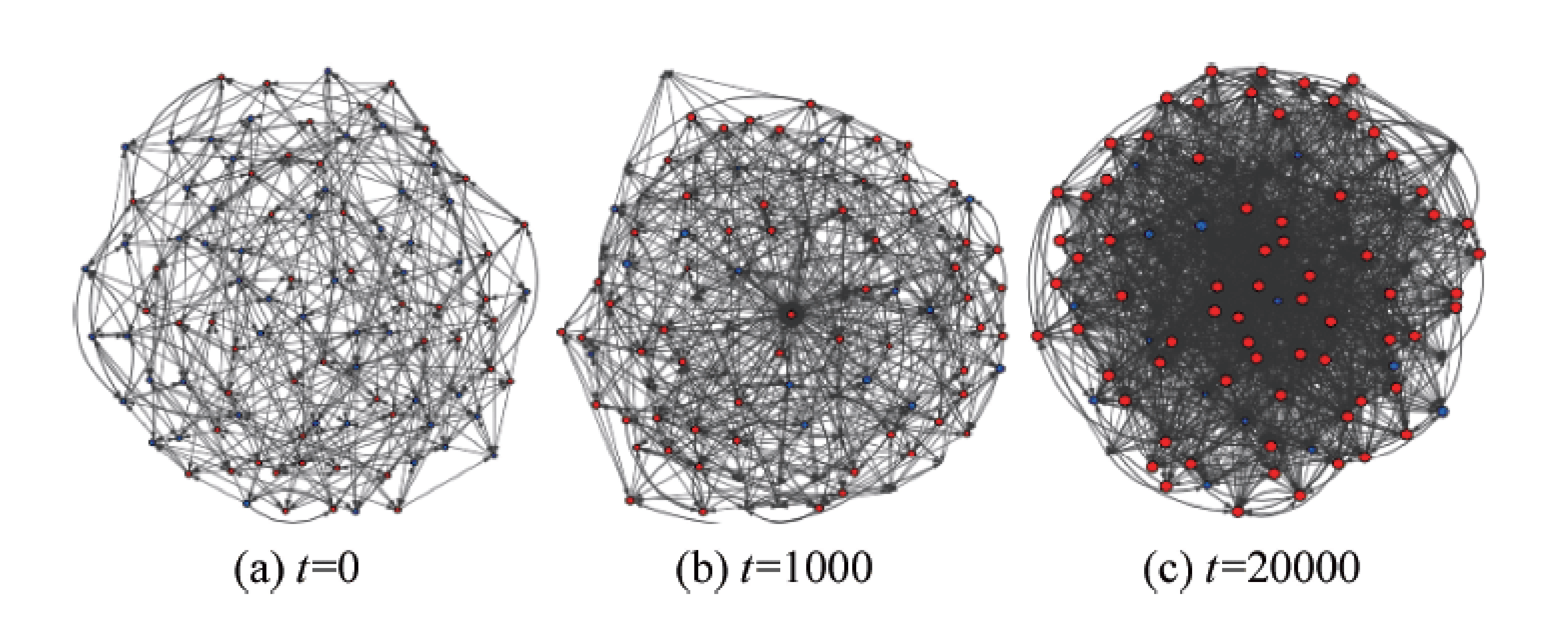


图 6-4　网络演化过程图

合作者标为红色，背叛者标为蓝色。图中个体的大小与个体的声誉值有关，取值是个体声誉值的对数。参数设置：$N=100$，$W=0.8$，$p_r=0.8$，$r=3.2$

信息分享行为的演化进行了分析。如前所述，合作者比例 f_c 作为评价指标来刻画用户间合作水平，它可以表示为重要参数(W,p_r,η)的方程形式。在真实世界中，这 3 个参数和评价指标并不容易获得，但是通过使用近似方法对新浪微博数据进行分析，可以近似估计得到 W 和 f_c 的值。当得到估计值 $\hat{W}$ 和 $\hat{f}_c$ 后，将它们代入到上节的数值仿真结果中以估计新浪微博用户对声誉的追逐程度 $\hat{p}_r$ 和群体放大效应 $\hat{\eta}$。为了能够估计出 $\hat{W}$ 和 $\hat{p}_r$，本文使用了两种不同的采集自新浪微博的数据集。

行为策略更新与关注关系更新的时长比例 W 可以转化为一段时间内关注关系更新次数与用户行为策略更新次数之比。关注关系更新次数由用户在一段时间内的关注关系变化的次数表示，而行为策略更新次数则可以由用户合作行为次数来近似表示，即以用户在一段时间内的发帖次数近似表示。因此，W 可以近似为关注关系变化数与发帖数之比。用于估计 $\hat{W}$ 的数据集是分别在 2013 年 2 月和 2014 年 9 月初采集的新浪微博的 124 587 个账户信息，内容包括用户 ID、关注数、发帖数等。在这段时间内，这些用户的关注数变化了 6 592 158 次，发帖 65 756 553 次。由此，$\hat{W}=0.100\ 251$。

由于用户的真实行为策略难以准确获得，以及数值仿真中的演化周期(evolutionary time step)也难以与现实情况进行准确的映射，合作者比例 $\hat{f}_c$ 也需要近似估计。如前所述，

用户的合作行为可近似为用户的发帖行为，这样，$\hat{f}_c$可以近似地由用户的活跃程度来表示。为了能够刻画用户的行为策略，将用户的平均活跃度设为“标准”活跃度，作为用户合作与否的判定阈值。当用户的活跃度高于标准活跃度时，则认为他是合作者，因为他的表现足够活跃。相对地，如果用户的活跃度低于标准活跃度，则认为他是不合作者。用于估计$\hat{f}_c$的数据集是新浪微博的 71 118 个用户在 2014 年 3 月 28 日～2014 年 3 月 31 日的发帖行为，总发帖数为 904 462。通过近似方法，合作者比例的估计值为$\hat{f}_c=0.108\,875$。

当把估计值 $\hat{W}$和$\hat{f}_c$代入 6.4 节的数值仿真结果中时，可以看到，最接近$\hat{f}_c$的仿真结果的f_c位于图 6-1(b)和图 6-2(b)中的$\hat{p}_r=0.9$ 和$\hat{\eta}=0.6$。较高的$\hat{p}_r$说明新浪微博的用户在建立新的关注关系时喜欢选择高声誉用户。而较低的$\hat{\eta}$则说明新浪微博中群体放大效应对于合作的促进不足，需要寻找恰当的机制来增强群体放大效应对用户效用的影响。

用户在社交网络中分享信息的行为被认为是一种群体行为，可以使用公共品博弈进行建模。对社交演化博弈模型的数值仿真揭示了关注关系更新频率、用户对声誉的追逐程度和群体放大效应对用户合作行为演化的影响。关注关系更新频率的增大会使社交网络中的用户难以达到高的合作水平，但是合作者会更容易产生和生存。用户对声誉的追逐程度和群体放大效应的增大能够促进用户的合作，但是合作水平会受到关注关系更新频率的制约。由此可以看到，相比对声誉的追逐程度和群体放大效应，关注关系更新频率对于合作水平具有两面性。除了数值仿真实验，本文应用社交演化博弈对新浪微博的用户信息分享行为的演化进行了分析。通过使用近似的方法，从真实数据集中估计得到$\hat{W}$和$\hat{f}_c$，并将它们代入数值仿真结果中来以估计用户对声誉的追逐程度和群体的放大效应。结果显示，新浪微博用户表现出对声誉的偏好，并且需要恰当的激励机制增强群体的放大效应以便促进合作。

6.5　信息交互行为对网络结构的影响

6.5.1　信息交互行为模型

信息交互行为的双向性使得社交网络个体之间的信息交互化简为无向交互，因此，社交网络可以被表示成一个无向网络 $G=(V,E,I;U,R)$，网络节点表示个体($|V|=n$)，边表示个体之间的社交关系($|E|=m$)，I 表示个体间的交互，而 $U=\{U_1,U_2,\cdots,U_n\}$ 和 $R=\{R_1,R_2,\cdots,R_n\}$则分别表示个体的效用函数和声誉函数。

网络中的个体都是自私的，他们只关注自身的效用，同时他们也是短视的并缺乏全局信息，只能获取到其所处环境的局部信息，即他们的邻居节点和邻居节点的邻居节点的信息。个体有两种交互策略$\{C, D\}$，策略 C 表示正向行为，使用该策略的个体在交互中会付出一定的成本 c，而其对手则会获得收益 b；策略 D 表示负向行为，使用该策略的个体在交互中不付出任何成本，他只是单纯地获取收益。个体与个体之间的交互是两两进行的，使用经典的囚徒困境模型进行刻画，个体 i 与个体 j 博弈时，i 的收益矩阵见表 6-1。为了方便模型与仿真结果的描述，本文使用付出收益比(c/b，cost-to-benefit ratio)对表 6-1 的收益矩阵进行变换，得到新的收益矩阵 A，见表 6-2。由此，个体 i 的效用是他与所有邻居节点进行博弈所获收益之和，即

$$u_i = \sum_{j \in N_i} A(s_i, s_j)$$

其中，N_i是 i 的邻居节点集。

表 6-1　收益矩阵（表中是 i 的收益）

i \ j	C	D
C	$b-c$	$-c$
D	b	0

表 6-2　变形后的收益矩阵（表中是 i 的收益）

i \ j	C	D
C	1	0
D	$1+c/b$	c/b

6.5.2　更新机制

社交演化博弈的演化规则表示了个体短视的决策过程。个体的短视是指由于网络的限制，个体只能获取到个体所处的局部环境的情况，具体来说就是他只能得到其邻居节点和邻居节点的邻居节点的信息。在每一时刻，网络中随机的一个个体进入更新状态。当一个个体进入更新状态时，他可以选择更新自己的社交关系，改善所处的声誉环境，或者选择模仿成功的邻居节点的行为策略来改善效用。具体地，当个体 i 进入更新状态时，他以概率 τ 进行社交关系更新；否则，进行策略更新。

基于个体 i 的当前声誉值r_i，将声誉划分为 3 个区间。

$$\begin{cases} r_i^+ = \{R \mid R > r_i + \lambda \\ r_i^\sim = \{R \mid r_i - \lambda \leqslant R \leqslant r_i + \lambda\} \\ r_i^- = \{R \mid R < r_i - \lambda\} \end{cases}$$

其中，λ 是个体的声誉容忍阈值。个体 i 有两种方式寻找新个体建立社交关系：①依概率p_r选择邻居节点的邻居节点中声誉值落入区间r_i^+ 的最大声誉值的个体；②否则，随机地从邻居节点的邻居节点中选择一个个体。这里，概率p_r反映了个体对于声誉的偏好程度。个体 i 在建立社交关系的同时，也会断开他不满意的社交关系，即断开与邻居节点中声誉值落入区间r_i^- 的最低声誉值的个体的社交关系。

当个体进行策略更新时，他模仿依概率 φ 邻居节点中效用最大的个体的策略，即

$$\begin{cases} \varphi = \dfrac{1}{1 + \mathrm{e}^{-\beta \Delta u}} \\ \Delta u = u_j - u_i \end{cases}$$

其中，β 表示个体模仿其他个体时的非理性因素。

当个体进行更新策略时，其声誉值也同步更新。

$$R_i(t) = R_i(t-1) + \Delta R$$

当 $s_i = 1$ 时，$\Delta R = 1$，否则 $\Delta R = 0$。

6.5.3　实验结果与分析

基于以上社交演化博弈模型，我们在付出收益比 c/b 和声誉的偏好度p_r的不同组合下对模型进行了仿真，每组参数下独立执行了 500 次仿真。初始时，网络结构是节点平均度

$\langle k\rangle=4$ 的随机网络，每个节点以相同的概率被赋予正向策略或负向策略。仿真结果采用 4 种指标进行评价：个体的正向策略密度 f_c、终态网络聚集系数 cc、终态网络度分布幂律指数 α 和社交演化时间。

当网络演化达到如下 3 个条件之一时，则认为网络演化达到相对稳定：全正向策略、全负向策略，或者网络中正向策略的比例在一段长时间内的变化率较小。

图 6-5(a)显示了演化终态时网络中使用正向策略个体的比例情况。

从图中可以看到，当付出收益比 c/b 接近于 0 时，终态网络中的所有个体都使用正向策略，当 c/b 大于一定阈值时，终态网络中的所有个体都使用负向策略，因此，p_r 对终态时的正向策略比例 f_c 的影响不大。当 c/b 落在一定区间内时，p_r 的变化对终态网络的正向策略比例影响较大，正向策略比例 f_c 随 p_r 的增加而增加。图 6-5(b)则显示了终态网络的聚集系数，c/b 在 0.1 附近时，p_r 的变化对终态网络的聚集系数 cc 影响明显，cc 随着 p_r 的增加而增加。与图 6-5(a)进行对比可以看到，当网络中个体更偏好声誉时，终态网络更加聚集。图 6-5(c)显示了度分布的幂律分布指数，与图 6-5(a)类似，c/b 和 p_r 也存在一定的区间使得它们的变化对 α 的影响较大。而图 6-5(d)则显示了演化达到相对稳定时所需要的演化时间。与图 6-5(a)进行对比可以看到，当网络无法维持正向策略时，网络较容易达到全负向策略，演化时间较短。而大多数演化时间较长的区间多出现在终态网络正向策略比例在 0.6 附近时。

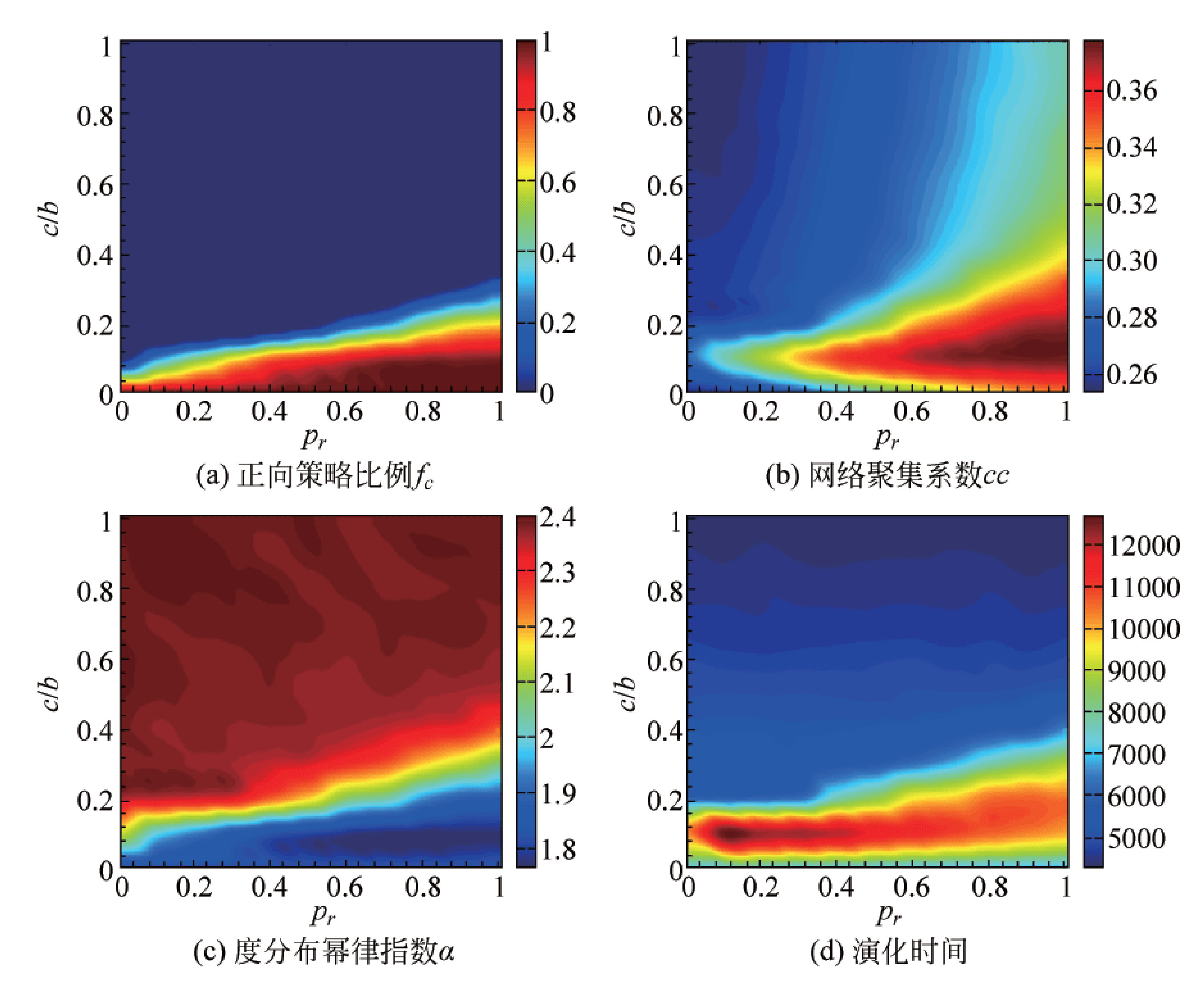


图 6-5　社交网络演化终态的合作者比例及不同参数 c/b 和 p_r 组合下的社交网络演化终态的合作者比例

参数设置：$N=500$，$\beta=0.04$，$c/b=0.1$，$\tau=0.8$

综合以上现象的分析可知，社交网络演化中的声誉机制能够促进网络中正向策略的产生，并且使得网络结构演化得更加紧密。与此同时，个体对于声誉的偏好使其能够承受更大的付出。本应用案例的实际数据分析部分见 12.1 节。

小结

本章研究了社交网络中个体信息行为与关系结构共同变化下个体信息行为演化及其对社交网络结构的影响。个体信息行为按个体间的交互方式可分为个体信息分享行为和个体信息交互行为。

由于个体信息分享行为所具有的方向性和群体性，本章将社交网络抽象为有向网络，并使用公共品博弈描述个体信息分享行为。每一个个体既主持由他及其邻居个体组成的公共品博弈，也参与由其邻居个体主持的公共品博弈。他的效用是其参加这些公共品博弈的收益之和。本章分析了个体有向关系更新频率、声誉偏好和群体放大效应对个体信息分享行为演化的影响，得到了演化规律图。

针对个体信息交互行为具有的双向交互特点，本章将社交网络抽象为无向网络，并使用囚徒困境博弈描述个体信息交互行为。每个个体与其邻居个体分别进行两两交互，而其效用则是这些两两交互的收益之和。通过合作者比例、社交网络的聚类系数、节点度分布的幂指数和演化稳定时间这 4 个指标，本章分析了个体关系更新频率、声誉对个体信息分享行为的演化性及其对社交网络结构演化的影响。

通过本章的研究与分析可以发现，社交网络中个体的信息行为，无论是信息交互行为还是信息分享行为，都会促使个体的信息行为和关系产生聚集。

参考文献

[1] Hirahara Y, Toriumi F, Sugawara T. Evolution of cooperation in meta-rewards games on networks of WS and BA models. In: Proc. of the 2013 IEEEAVIC/ACM Int'l Joint Conf. on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Vol. 3. 2013. 126-130.

[2] Zimmermann MG, Eguíluz V. Cooperation, social networks and the emergence of leadership in a prisoner's dilemma with local interactions. Phys. Rev. E, 2005, 72: 056118.

[3] Equíluz VM, Zimmermann MG, Cela-Conde CJ, *et al*. Cooperation and the emergence of role differentiation in the dynamics of social networks. Am. J. Sociol, 2005, 110: 977-1008.

[4] Fu F, Hauert C, Nowak MA, *et al*. Reputation-Based partner choice promotes cooperation in social networks. Physical Review E, 2008, 78(2): 026117.

[5] Barabasi A-L, Albert R. Emergence of scaling in random networks. Science, 1999, 286(5439): 509-512.

[6] Pacheco JM, Traulsen A, Nowak MA. Coevolution of strategy and structure in complex networks with dynamical linking. Physical Review Letters, 2006, 97(25): 258103.

[7] Pacheco JM, Traulsen A, Nowak MA. Active linking in evolutionary games. Journal of Theoretical Biology, 2006, 243(3): 437-443.

[8] Amaral LAN, Scala A, Barthelemy M, *et al*. Classes of small-world networks. Proc. of the National Academy of Sciences, 2000, 97(21): 11149-11152.

第 7 章
群体行为与网络结构的协同演化

7.1 概述

由前述第 5 章和第 6 章可以看出，网络结构对网络上的信息传播产生影响，而用户的信息行为也能够对网络结构的演化产生影响。本章将关注用户聚集产生的群体的行为与网络结构之间的协调演化。

网络用户通过各种信息行为间的作用关系有机地关联在一起，形成具有一定拓扑结构关系的网络群体。一般地，网络群体是指互联网上有着相似特质、共同目标或价值观以及团体意识的，两个或两个以上的用户，以一定方式进行线上线下交互或活动而构成的用户群体。网络群体不但具有一般群体的特点，如网络群体成员有着相似的特质、共同的目标或价值观；成员对群体有认同感和归属感；群体成员间建有关系，构成群体的结构等，还具有其自身的特点：依托于社会关系网络，网络群体不受地理等因素的影响，成员分布范围极广，网络群体的时效性比一般群体强等。网络群体形成的方式多种多样，大致归类为以下几种：①由某种固定关系形成的群体，通常是现实社会中的实际关系在网络中的具体映射，例如由同学及校友关系形成的网络校友群体，这类群体的结构比较稳定，形成之后随时间的变化不大，成员的活跃程度有限；②由某种不确定的共同需求形成的群体，例如，微博中的微群、豆瓣的群组等，这类群组中的成员多是由共同的需求或目的聚集在一起，随着时间的变化，群体成员会发生一定的流动，但是流动的主要是度数相对较低的节点；③由某种突发事件驱动而形成的群体，这类群体具有很强的时效性，通常由一个焦点事件引发产生，其结构变化与该事件的关注度涨落有关，在事件平息后的短时间内群体就会解散。

在社交网络中，微观层面的个体信息行为在宏观层面的网络群体行为中所表现出来的特征更加复杂。网络群体的行为受到社会性和随机性等的影响，这使得群体的行为往往表现出极大的不确定性，而更加值得关注的现象是群体行为的涌现现象。由用户行为导致的信息的传播一般都具有从逐渐积累到一夕爆发的特点，这样的积累并不是线性

的，也不是用户行为产生效应的简单迭加，而是一系列小的局部变化，每一个小的局部变化都不足以对整体系统造成影响，但当这些变化的积累达到某个临界状态时，整个系统出现临界相变。整个系统表现出来的现象是所有局部行为相互作用的结果。社会关系网络中的大量涌现现象包括信息扩散数量上的涌现、网络结构特征的涌现、特定网络群体的涌现以及特定网络群体行为的涌现等。

作为理论简化和直观表达，对网络群体行为进行分析和建模是社交网络研究的重要内容。目前，对于网络群体行为的模型的研究集中在对网络群体行为定性描述的模型上，主要作为表示和呈现个体信息行为复杂过程的一种方法。社会关系网络是一个多维度的复杂系统，网络内部节点之间的相互影响与相互作用频繁。然而，由于网络群体行为的社会性、随机性和复杂性使得传统的统计研究方法不能对网络群体行为产生的影响动态走向进行预测，因此迫切需要新的基础理论和研究方法。目前，网络群体行为模型的相关研究工作主要包括信息传播及观点交互模型、群体行为演化模型、群体结构模型和突发群体事件模型等，但还没有一个公认的成体系的研究思路。近年来，对社交网络的建模方面进行了大量研究工作，比如对于社会网络静态社区结构特性的研究、在演化过程中社会网络动态社区结构变化规律的发现等。但是，这些研究工作的关注点主要集中在网络拓扑结构上，对于网络群体行为的预测调控工作超出了其研究范围。

针对社交网络群体的研究，本章首先介绍了可用于网络群体评价的指标，包括群体结构评价指标和群体行为评价指标。随后介绍了信息交互群体识别方法，在社交网络中，用户间频繁的信息交互行为促使他们的联系更加紧密，形成信息交互群体，而信息交互群体则对用户间的信息交互产生促进作用，使群体成员的交互更加紧密。本章提出的信息交互群体识别方法可以有效地识别出信息交互群体。而针对跨网络的重叠群体的行为与结构的协调演化，本章也进行了初步的探索。此外，针对跨网络群体的用户关注度竞争问题，本章提出了一种竞争性社交演化博弈模型用以描述在一个群体之上的多个社交网络用户关注度的竞争过程。

7.2　社交网络群体评价指标

为了能够定量地分析网络群体行为，有必要建立可量化的评价指标体系。现有的工作主要集中在使用过程模型对网络用户信息行为进行定性描述，以及对个体行为的研究。文献[1]对网络中用户个体行为的可信性进行了研究，提出了统一度量的框架。而网络群体定量分析的研究还处在起步阶段，网络群体行为的量化评估并没有系统的规划，也没有完整的指标体系。

由于网络个体行为对于网络群体行为的涌现趋势和演化结果有着重要的影响，对个体信息行为进行分析评估和预测也是网络群体行为研究中的重要内容。本文将网络群体行为的评价指标分为 3 类：网络个体行为的评价指标、网络群体行为的评价指标和网络群体之间行为的评价指标。这 3 类指标分别从个体、群体和群体间 3 个层面综述了网络群体行为的评价指标。这些指标反映了个体和群体不同侧面的特点，在对网络群体进行分析评价时，所选择的评价指标依赖于待解决的问题及应用场景。本章总结了可用于刻画网络群体行为的评价指标。如图 7-1 所示。

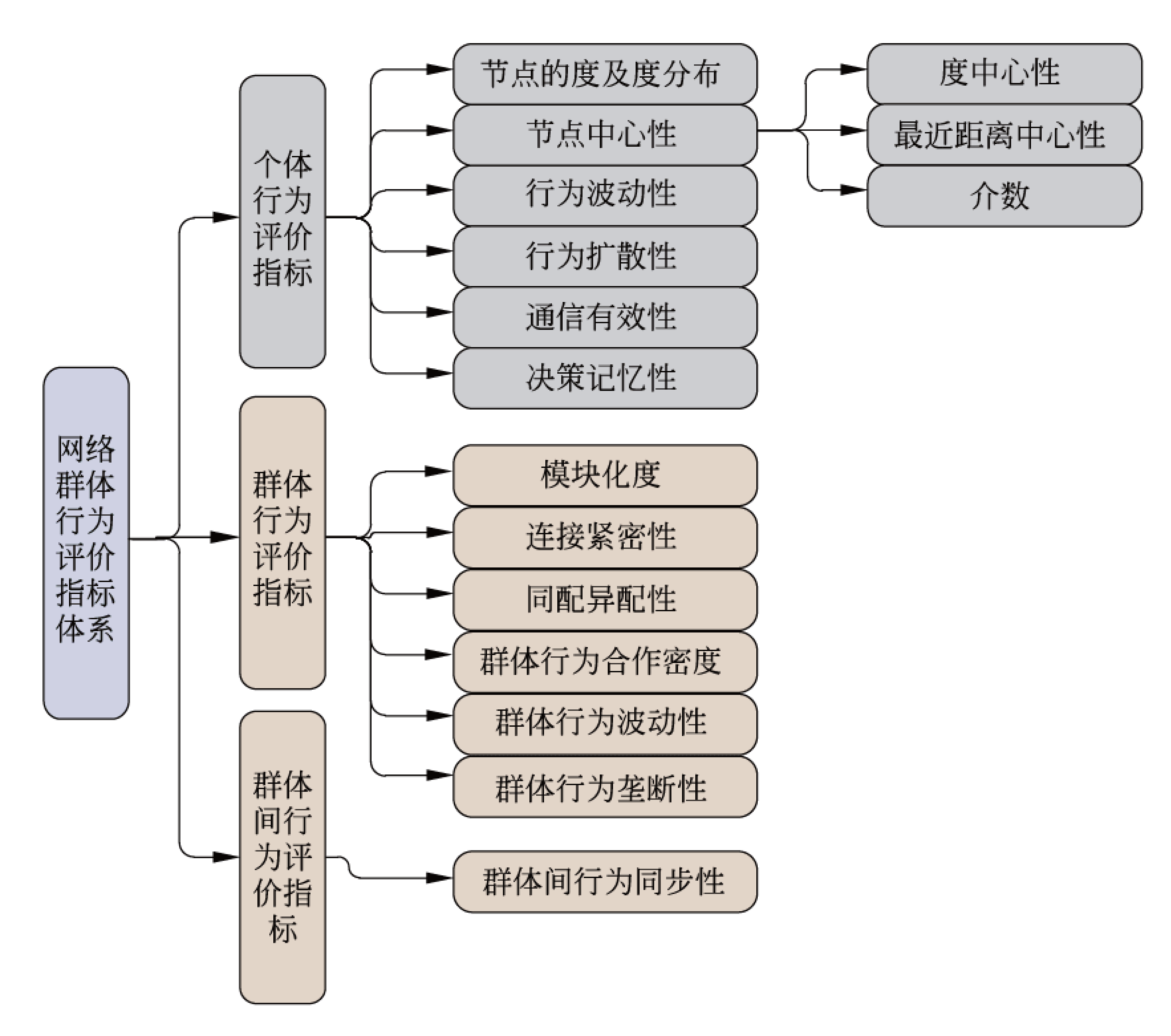


图 7-1　网络群体行为评价指标

7.2.1　群体结构评价指标

(1) 连接紧密性

连接紧密性通常用来衡量网络联结的紧密程度，类似于社会关系网络中"物以类聚，人以群分"的特性，即考察连接在一起的集团各自的近邻之中有多少是共同的近邻。常用的指标为聚类系数(clustering coefficient)，它表示网络中节点聚集成团的倾向，可以细分为全局聚类系数、局部聚类系数和平均聚类系数。在社会关系网络分析中，常用的是局部聚类系数和平均聚类系数，在本文中如未特别指出，聚类系数均代指局部聚类系数。

节点v_i的聚类系数C_i是节点v_i的所有邻居节点之间实际存在的边数和这些邻居节点所有可能的边数之比。假设在一个图$G=(V,E)$中，V表示节点集，E表示边集，边e_{ij}连接节点v_i和v_j。设$N_i=\{v_j:e_{ij}\in E\wedge e_{ji}\in E\}$是节点$v_i$的邻居节点，$k_i=|N_i|$是$N_i$中节点的数目。对于有向图而言，边具有方向($e_{ij}\neq e_{ji}$)，$v_i$的聚类系数为

$$C_i=\frac{|\{e_{jk}:v_j,v_k\in N_i,e_{jk}\in E\}|}{k_i(k_i-1)}$$

而对于无向图，边不具有方向($e_{ij}\neq e_{ji}$)，则v_i的聚类系数为

$$C_i=\frac{2|\{e_{jk}:v_j,v_k\in N_i,e_{jk}\in E\}|}{k_i(k_i-1)}$$

整个网络的平均聚类系数C为所有节点的聚类系数的平均值。文献[12]将聚类系数推广到带权网络中。

(2) 同配异配性

同配异配性用来描述高度数节点连接的倾向性，也称作同配系数，其数学表达如下：

$$r=\frac{M^{-1}\sum_i j_i k_i-\left(M^{-1}\sum_i \frac{j_i+k_i}{2}\right)^2}{M^{-1}\sum_i \frac{j_i^2+k_i^2}{2}-\left(M^{-1}\sum_i \frac{j_i+k_i}{2}\right)^2}$$

其中，j_i 和 k_i 是第 i 条边连接的两个节点的度数，M 为总边数。正度相关性表示高度数节点倾向于与高度数节点连接，数值越大，倾向性越明显，这种现象在社会网络中比较常见；负度相关性表示高度数节点倾向于与低度数节点连接，绝对值越大，倾向性越明显，这种情况在生物网络中得到了很好的体现[3]。

(3) 模块化度

模块化度(modularity)是度量网络社区结构的重要参数，它同样可以用来分析和度量网络群体交互行为的结构特性。模块化度由 Newman 教授提出，探讨的对象主要是小世界网络和 BA 无标度网络两类典型复杂网络的模块化特点，其聚团性定义为

$$Q=\frac{1}{2m}\sum_s(e_{ss}-a_s^2)$$

其中，e_{ss}表示第 s 个社区内部边的权重之和占网络中边权重总和的比例，$a_s=\sum_t e_{st}$，$2m=\sum_s a_s=\sum_{st} e_{st}$ 是网络中边的总权重。上面的式子给出的模块度定义是定义在单个社区上的，即 $e_{ss}-a_s^2$ 对应着第 s 个社区对 Q 的贡献。模块度隐式地给出了社区的定义：$e_{ss}-a_s^2$，即社区内部边的权重之和比期望的边权重之和要高。

模块度的提出推动了网络社区结构的研究，它为研究人员提供了用于选择网络划分的目标函数，模块度的优化算法很快成为社区发现算法的主流。模块度在无权无向简单网络的社区发现上所取得的成功激发了研究人员把模块度扩展到其他类型的网络上，包括带权网络[4]、有向网络[5-7]、二部图[7,8]、具有负值边权的网络[9,10]等。使用不同的 null model，模块度可进一步细化，从而得到有参数的模块度[11,12]和基于 motif 的模块度定义[13]。

但是，网络的所有可能的划分数量巨大，Newman 指出，网络的所有划分个数是第 2 类 Stirling 数[4]，因此，在由网络划分构成的空间里寻找模块度最优的划分是困难的。Brandes 等人证明了模块度的优化问题属于 NP 完全问题[14]，给出了模块度优化问题的时间复杂性。

7.2.2　群体行为评价指标

网络群体行为评价指标主要评价的对象是网络群体行为特性以及群体结构特性。

(1) 合作密度

合作密度又称作合作水平，$\delta_x(t)$等于 t 时刻群体 G_x 中合作者的数量除以 G_x 的总个体数，也可称为合作频率，即在所有个体的总博弈次数中采取合作策略的比例[15]。

(2) 群体行为波动性

群体行为波动性的衡量指标为波动系数，它是所有个体改变行为选择的平均频率；波动(扰动)系数为 1 时，意味着所有个体在每一时刻改变他们的选择；该系数为 0 时，则意味着每个主体在进行了他们的初始选择之后不再改变[16]。

(3) 群体行为垄断性

群体行为垄断性用于描述群体行为的从众程度。设个体的行为选择集合为 $i\in\{0,1\}$，

该指数为 0 时，表示个体中选择 $i=0$ 和选择 $i=1$ 的数目各占一半，而当该指数为 1 时，表示所有个体或全部选择 $i=0$，或全部选择 $i=1$。该垄断指标由最后十步模拟的平均基尼系数算出。基尼系数的计算公式为

$$g=\frac{\sum_{i=0,1}\sum_{j=0,1}|y_i-y_j|}{2(|A|-1)\sum_{i=0,1}y_i}$$

其中，y_i 代表选择行为 i 的数目，$|A|$ 表示集合中元素的个数[16]。

群体行为评价指标侧重中观层面从群体的角度来分析和评估群体行为的演化结果和影响。这些指标中，连接紧密性和模块化度反映了群体结构特性，即节点间的连接关系特性；通配异配性和合作密度群体行为垄断性反映了群体行为和个体的倾向性，其中，通配异配性反映了大度数节点的连接关系倾向性，合作密度反映了群体内个体采取合作行为的水平，而群体行为垄断性反映了群体内个体行为的倾向程度；群体行为波动性则反映了群体内个体行为的变化情况，即群体内部行为的稳定状况。从上述可以看出，根据待解决问题的侧重点和应用场景，选择不同的评价指标来分析评价群体中的个体及其行为，如反映群体结构特性的连接紧密性和模块化与反映群体行为和个体倾向性的通配异配性、合作密度和群体行为垄断性搭配组合时，可以刻画群体结构与群体行为的倾向、群体采取的行为等之间的关系。当反映群体行为和个体倾向性的指标与群体行为波动性搭配组合时，则可以刻画群体或个体的倾向性与群体行为变化之间的关系等。

7.3 社交网络群体识别方法

在线社交网络服务平台和应用能够使不同地域的人们通过互联网联系起来并建立社会关系，满足人们的社交需求。这使得众多的在线社交网络服务平台和应用得以广泛地流行，并拥有了大量的用户，例如 Twitter、Facebook、新浪微博等，都拥有上亿的用户。

社交网络中的好友关系、关注关系等由于变化缓慢，在一段时期内可以看作是“静态的”，相比而言，用户之间的交互则频繁得多。在社交网络结构保持“静态”的一段时间内，随着时间会有大量的交互产生，因此，交互可以看作是“动态的”。相比社交关系，交互通常由于带有属性而包含了更多的信息。例如，用户间频繁的交互可能意味着他们之间的关系非常紧密。

近年来的一些实证研究也倾向于分析社交网络中的交互而非关系。与社交网络中社交关系构成的社交结构类似，由交互构成的网络也存在聚集群体现象。但是，由于交互与社交关系的不同特点，这种交互的聚集群体与网络结构中的社区是不同的[1]。社交网络中存在多种类型的交互。有些交互可能只是简单的一个用户行为，如点击网页上的“喜欢”按钮，而有些交互要复杂一些并涉及内容，如用户的发帖和回复行为。在本章中，关注并探讨后者。

本章中，交互是指用户间对于信息交流的成对的序列化行为。举例来说，假设有一个用户在 Twitter 或 Facebook 上发布或评论了一条消息，接收到这条消息的用户则可能对其进行评论、回复或转发。在发布者和接收者之间的这种成对的序列化行为就构成了一个交互。由于带有内容的交互蕴含着社交信息，故而可以将其看作是一种强的社交连结。这种强的社交连结所蕴含的社交信息并不能被社交网络结构或用户描述所反映。

下面给出了几个例子来说明本章的关注点。

- 在社交网站中，如 Facebook，用户可以在其主页或加入的群组中发布消息、图片等。在这种情形下，用户可以通过发布状态、相互回复的方式来相互交流彼此的喜好或关注的事物。因此，可以通过用户的“发帖-回复”行为为连结构建用户交互行为的社交网络。明显地，“发帖-回复”行为附带的内容信息可以作为区分聚集群体的依据，以此识别出喜好或关注的事物相同的聚集群体。
- 在微博类应用中，如 Twitter，用户可以回复、转发消息或者在消息中“@”某用户来交互。在这种情形下，微博消息内容相似且存在交互的用户比微博消息内容相似且没有进行交互的用户更容易属于同一群体，即使这些没有交互的用户明确地以“#”进行了话题归类。

交互内容显示了参与交互的用户真正关注的事物。用户与不同群体内的用户交流的内容有可能是不同的。换句话说，用户的生活是多方面的，这导致了他会因为不同的关注事物与不同的用户进行交流。这种现象在现实生活中随处可见，在社交网络中也是如此。进一步地，当涉及交互群体时，交互的时间间隔也是一个重要的参数。如果一个交互与其他的交互间隔时间过长，或者构成这个交互的用户行为发生的时间过长，那么这个交互在识别聚集群体时就会失效。当把交互的时效性考虑在内时，有着相似内容的交互可以被看作是强连接。这种强连接体现了用户间的有效关联性。

7.3.1　相关研究工作

与社交网络中信息交互群体相类似，在真实世界的网络中，社区结构（以下称为社区）是一个重要的特征并且广泛存在。通常，社区被认为是网络中的一组具有相同属性和（或）相似角色的节点集合[17,18]。根据进行社区发现时所用到的信息，现有的社区发现方法主要可以分为 3 类：连接分析法、内容分析法和连接-内容组合方法。

基于连接的社区发现方法只使用了网络结构信息。这里，方法可以分为两类：基于度量的方法和基于模型的方法。基于度量的方法的目标是优化社区划分的指标，这类指标用来评估社区划分的质量。在这类指标中，一个具有里程碑式的指标是 Newman 等人提出的模块化度[19]。当使用该指标度量社区划分质量时，若得到的值最大，则所得到的社区划分被认为是最好的，或者至少是一个好的划分。尽管模块化度存在一些缺点，比如分辨率问题[20]，但它仍然被广泛采用。基于模型的方法则是建立在统计推断的基础上，比如产生式模型（generative model）[21,22]、block models[23]等。

基于内容的社区发现方法通常是将节点或连边的附加信息分为不同的话题，而每个话题都被认为是一个社区。Zhou 等人提出了一种名为“社区-用户-话题（community-user-topic，简称 CUT）”的概率模型[24]，该模型利用节点间的交流内容来发现语义社区。这类社区是由有着相似交流兴趣或话题的节点组成。基于内容的社区发现方法忽略了社交网络中的结构信息。

在社交网络中，比如 Facebook 和 Twitter，不仅有着网络结构的信息，还包含着附着在节点和连边上的内容信息。因此，当在社交网络上进行社区发现时，只使用连边信息或内容信息是不充分的。近年来，出现了一些整合连边信息和内容的方法。“社区-发布者-接收者-话题（community-author-recipient-topic，简称 CART）”模型[25]组合了连边信息和交流内容

来发现社区。该模型认为，在一个社区中的用户由于共同的兴趣而有着相同的话题。由于考虑了每条信息的所有接收者，导致了 CART 在面对具有广播特性的社交网络时缺乏实用性。Sachan 等人[26]提出了“话题-用户-接收者-社区(topic-user-recipient-community，简称 TURCM)”模型，该模型用到了话题、网络结构和交互来发现社区。该模型克服了 CART 的缺陷，在安然邮件网络(Enron Email network)和 Twitter 上都表现很好。除了产生式模型，Qi 等人[27]则使用了一种名为 edge-induced martix factorization(EIMF)的矩阵分析方法。该方法将连边信息嵌入到基于网络连边结构的隐向量空间中，使有着相似内容的连边的隐向量被聚集到一起，从而得到社区划分。

由于社区的出现是由社交网络中个体的行为导致的，并且这些个体行为都各自发生，没有一个中心的力量来达到一个全局目标，因此博弈模型方法也被引入到社区发现领域。作为早期的一个尝试，Chen 等人[28]提出了一种基于博弈理论的方法以识别真实网络中的重叠现象[29]。该博弈模型方法使用了一种派生自 Newman 的模块化度的个体化模块化度作为个体的收益。在 Alvari 等人[30]提出的社区形成博弈模型中，个体的效用则是基于等结构(structure equivalence)这一定义。这两种算法都是随机算法，需要大量的重复计算来降低随机性的影响。因此，Narayanam 和 Narahari[31]提出了一种基于确定性算法的社区发现博弈模型，该模型中每个个体的效用基于他的邻居的稠密程度。Basu 和 Maulik[32]利用个体的选择偏好提出了特征联盟博弈模型来发现稳定致密的群体结构，并通过使用贪婪算法来计算纳什均衡。Jiang 和 Xu[33]使用参与程度和友好强度作为个体加入一个群体的收获函数，而损失函数则分为在当前群体的缺陷和与群体外的关系强度两部分，他们提出的方法只能发现非重叠群体。Zhan 等人[34]将网络划分为无重叠部分和重叠部分，并使用了两种测量方法来计算一个节点的信息量，提出了刻画无重叠部分和重叠部分共同演化的演化博弈方法，从而用以发现网络的重叠群体。

这些基于博弈理论的方法都只使用了结构信息。如前所述，信息交互在发现社区或群体中扮演着重要的角色。因此，本章根据 Chen 等人[28]提出的方法，基于博弈理论和信息交互来识别信息交互群体。在本文方法中，使用交互及其内容而不是连边来识别社交网络中的信息交互群体。

本文的基于博弈论的识别信息交互群体的方法本质上是自私的个体在交互网络上进行的策略博弈。每个个体的效用和与其进行交互的个体有关。在博弈中，个体根据对手的策略从自身的可选策略中选择最佳的策略以最优化自身的效用。信息交互群体是这种网络个体自私交互所导致的结果。在对本文进行详细介绍之前，我们先来介绍要用到的一些基本概念。

7.3.2　信息交互群体识别方法

定义 7.1(交互网络)　交互网络是由个体及带有附属信息的个体交互组成的网络。设 $\mathcal{J}=(V,E;\Gamma,\Delta T)$ 表示一个交互网络，其中，V 是个体集合；E 是个体的交互集合；Γ，ΔT 则分别是 E 中交互的附属信息——交互内容和时间间隔的相应集合。

设 I 是交互集合 E 中的一个交互，$\Delta t_I \in \Delta T$ 是 I 的时间间隔，$\gamma_I \in \Gamma$ 是 I 的内容。Z 是从 Γ 中产生的话题向量，Δt 是预设的交互时间间隔的阈值，Δ_{Sim} 则是交互内容与话题的相似度阈值。

定义 7.2（有效交互）　设 I 是一个有效交互，则它满足 $\Delta t_I \leqslant \Delta t$ 并且 $\mathrm{Sim}(\gamma_i, Z) \geqslant \Delta_{\mathrm{Sim}}$，其中，$\mathrm{Sim}(\gamma_I, Z)$ 是 γ_I 和 Z 的函数，返回 γ_I 与话题向量 Z 中最大的相似度。

定义 7.3（有效交互网络）　设 $\mathcal{J}=(V', E'; \Delta t, \Delta_{\mathrm{Sim}}, Z)$ 是一个有效交互网络，则它满足 $V' \subseteq V, E'=\{I \mid \forall I, I \in E, \Delta t_I \leqslant \Delta t, \mathrm{Sim}(\gamma_I, Z) \geqslant \Delta_{\mathrm{Sim}}\}$。

简单地说，有效交互网络是个体及个体间的有效交互构成的网络。

定义 7.4（信息交互群体）　设 $\mathcal{G}=(V_g, E_g, z)$ 是一个信息交互群体，其中，$V_g \subseteq V'$ 是群体的成员集合，$E_g \subseteq E'$ 是群体的成员交互集合，$z \in Z$ 是群体的话题。

信息交互群体是有效交互网络的子集。每个信息交互都有两个属性，即内容和时间间隔。一个信息交互的内容由一个发布者和一个接收者最初发布的消息及后续一系列回复的消息组成，而这个信息交互的时间间隔则是第 1 个消息的发布时间与最后一个消息的发布时间的时间差。信息交互群体就是在这样的信息交互构成的信息交互网络上进行识别的。信息交互网络的动态性体现在以下两个约束上，即群体共享的话题和信息交互的时效。基于这两个约束，在进行信息交互群体识别时，需要对每一个信息交互进行有效性检验。通过移除非有效信息交互，可以从信息交互网络中得到有效信息交互网络，由此可见，有效信息交互网络是信息交互网络的子集。信息交互群体的识别就是在经过化简之后的有效信息交互网络上进行的。

因此，信息交互群体识别方法分为两个阶段：化简阶段和识别阶段。在化简阶段中，通过对信息交互的有效性检验以对信息交互网络进行化简，生成有效的信息交互网络；在识别阶段，使用在生成的有效信息交互网络上执行博弈模型方法从而进行信息交互群体的识别。这两个阶段均依赖于信息交互的内容。

首先，构建信息交互网络。网络上的每个节点都是自私的、短视的个体。每个个体的目标是根据自身感兴趣的话题选择参与或退出不同的信息交互群体来优化自身的效用。话题是从所有信息交互的内容产生的，并且每个群体使用一个话题进行标识。个体根据自身的感兴趣的话题可能属于一个或多个信息交互群体。信息交互群体构建完成之后，进入化简阶段。在此阶段，检验所有信息交互的有效性，删减非有效信息交互。此阶段结束之后，从信息交互网络得到有效信息交互网络。

一旦化简阶段结束，即进入识别阶段。在此阶段，个体将在有效信息交互网络上执行策略博弈。在此策略博弈中，个体的效用由两部分组成：收获和损失。收获反映了个体根据信息交互选择的信息交互群体相比随机选择的信息交互群体的优势，而损失则反映了个体为维持在信息交互群体内所付出的代价。

当策略博弈进行到没有个体能够改善收益时，策略博弈达到均衡状态，即所有的个体均满足他们所选择的信息交互群体。由于个体缺乏全局知识而无法知晓网络上所有的话题，因此与文献[28]类似，本文也使用局部均衡（local equilibria）[35]代替全局均衡作为策略博弈的解。

1. 个体交互博弈模型

本节详细介绍个体交互博弈模型，该策略博弈的思想继承自 Chen 等人的工作[28]。

给定一个信息交互网络$\mathcal{I}=(V,E;\Gamma,\Delta T)$，其中，$|V|=n$，$|E|=m$。信息交互的内容涉及的话题使用话题模型生成(如，Latent Dirichlet Allocation，LDA 模型)从内容向量Γ中产生。设$Z=\{z_1,z_2,\cdots,z_k\}$是从Γ学习产生的话题集。每一个可能的信息交互群体由一个话题标识。当化简阶段结束，得到有效信息交互网络$\mathcal{I}=(V',E';\Delta t,\Delta_{\text{Sim}})$，其中，$|V'|=n'$，$|E'|=m'$。博弈模型方法在$\mathcal{I}$上执行。每个个体的策略是其所属的话题集。设$Z_i$是个体$a_i$参与过的话题的集合，$|Z_i|=k'$是其参与过的话题个数，$s_i\subseteq Z_i$是$a_i$的一个具体策略，即其所属的话题，$a_i$的策略空间$S_i$是$Z_i$的幂集。

如果a_i使用的策略是空集，即$s_i=\varnothing$，则意味着a_i不属于任何话题，即不属于任何信息交互群体，但是这并不意味着他没有参与过任何话题。

设$\{s_1,s_2,\cdots,s_{n'}\}$是$I'$中所有个体的一个策略组合，而$s_{-i}$是除了个体$a_i$之外所有个体的策略组合，则当$a_i$使用的策略是$s_i$时，所有个体的策略组合可表示为$(s_{-i},s_i)$。$a_i$的效用由收获函数$g_i(\cdot)$和损失函数$l_i(\cdot)$构成，具体地，

$$u_i(s_{-i},s_i)=g_i(s_{-i},s_i)-l_i(s_{-i},s_i)$$

这里，收获函数使用了 Newman 模块化度的一个扩展版本[36]，该版本被认为是更适合度量存在重叠社区的网络。

$$g_i(s_{-i},s_i)=\begin{cases}\dfrac{1}{2m'}\sum\limits_{1\leqslant j\leqslant n'}\dfrac{1}{|s_i|\cdot|s_j|}\left(A_{ij}\hat{\delta}(ij)-\dfrac{d_id_j}{2m'}|s_i\cap s_j|\right), & |s_i|\geqslant 1,|s_j|\geqslant 1\\0, & |s_i|=0\end{cases}$$

其中，A是网络$\mathcal{I}$的邻接矩阵，A_{ij}是A的元素；d_i和d_j分别是a_i和a_j在网络$\mathcal{I}$中的度；如果$|s_i\cap s_j|\geqslant 1$，则$\hat{\delta}(i,j)=1$，否则，$\hat{\delta}(i,j)=0$。个体的损失函数反映了个体在信息交互群体内维持与其他成员交流的成本。

由于每个个体至少会参与过一个话题，因此一个个体至少应当属于一个信息交互群体。这与文献[28]中的损失函数是类似的。不同的是，在这里，如果一个个体最终没有加入任何信息交互群体，那么他仍然要为已发生的交流付出成本c。

$$l_i(s_{-i},s_i)=\begin{cases}(|s_i|-1)\cdot c, & |s_i|\geqslant 1\\c', & |s_i|=0\end{cases}$$

其中，$c>0$是一个常量，个体a_i为其没有加入任何信息交互群体付出的成本是c'，为了简单起见，$c'=c$。

2. 局部均衡的存在性分析

在本文的策略博弈中，纳什均衡意味着当其他个体的策略保持不变时，没有个体能够通过改变自己的策略而改进收益。换句话说，每个个体都满足于其所属的信息交互群体。由于每个个体可能会对多个话题感兴趣而参与多个话题，因此，在纳什均衡点所得到的解会自然地包含这种情况。

定义 7.5(纳什均衡)　给定一个有效信息交互网络$\mathcal{I}$。S表示所有个体的全局策略空间，它实际上是话题集Z的幂集。如果没有任何一个个体能够通过单方面改变策略来改进收益，则所有个体的策略组合$\{s_i|s_i\in S,1\leqslant i\leqslant n'\}$是一个纳什均衡点，即$\forall i,s_i\neq s_{i'},s_i\in S,s_{i'}\in S$：$u_i(s_{-i},s_i)\geqslant u_i(s_{-i},s_{i'})$。

一般地，并不是所有的策略博弈都存在纳什均衡。然后，对于具有有限个策略组合的势博弈（potential game）来说，纳什均衡总是存在的[37]。具体地，在一个势博弈中，势函数 $\Phi(\cdot)$ 的定义与策略组合相关，即

$$\Phi(s_{-i},s_i)-\Phi(s_{-i},s_{i'})=u_i(s_{-i},s_{i'})-u_i(s_{-i},s_i)$$

当任一个体 a_i 改变策略来改进自身收益时，势函数减少的值与 a_i 效用增加的值一样。为了保证本文的策略博弈存在纳什均衡解，需要先证明它是势博弈。

设 $\{g_i(\cdot)\mid 1\leqslant i\leqslant n'\}$ 和 $\{l_i(\cdot)\mid 1\leqslant i\leqslant n'\}$ 分别是 $\mathcal{J}$ 中所有个体的收获函数和损失函数的集合，有 $g(\cdot)=\sum_{1\leqslant i\leqslant n'} g_i(\cdot)$ 和 $l(\cdot)=\sum_{1\leqslant i\leqslant n'} l_i(\cdot)$。如果收获函数和损失函数都是局部线性的，则可保证本文的策略博弈是势博弈。实际上，通过简单的数学推导即可验证收获函数和损失函数分别是线性因子为 $\frac{1}{2}$ 和 1 的局部线性函数。

本文策略博弈模型的势函数设为 $\Phi(s_{-i},s_i)=\rho l(s_{-i},s_i)$，$\Phi(s_{-i},s_i)=\rho_l l(s_{-i},s_i)-\rho_g l(s_{-i},s_i)$，其中，$\rho_l=1$，$\rho_g=\frac{1}{2}$ 是线性因子。根据局部线性函数和效用 $u_i(\cdot)$ 的定义，可以很容易地推导出上述公式。由此可知，本文的策略博弈是势博弈。

由于个体 a_i 参与的话题 Z_i 比全局话题 Z 要小，这导致了 a_i 的策略空间也比全局策略空间 2^Z 要小。S_i 是 a_i 的局部策略空间。在本文的策略博弈中，由于个体缺乏全局信息和自身的兴趣，他们都是自私短视的。因此，计算局部均衡比计算全局纳什均衡更合理，也更有效率。与纳什均衡类似，在局部均衡里，没有个体能够在自己的局部策略空间里通过改变策略来改进收益。

定义 7.6（局部均衡）　给定一个有效信息交互网络 $\mathcal{J}$，策略组合 $\{s_1,s_2,\cdots,s_{n'}\}$ 构成局部均衡，如果没有任一个体能够在自己的局部策略空间里通过单方面改变自己的策略而改进收益：$\forall i, s_i\neq s_{i'}, s_i\in S_i, s_{i'}\in S_i$，则 $u_i(s_{-i},s_i)\geqslant u_i(s_{-i},s_{i'})$。

本文的博弈模型方法使用局部均衡解代替全局纳什均衡解作为信息交互群体的识别结果。在 7.3.3 节中，将讨论计算局部均衡的复杂度，并给出识别方法两阶段中用到的主要算法。

7.3.3　局部均衡的计算

如前所述，本文的信息交互群体识别方法由两阶段组成，这两阶段都依赖于从信息交互中学习到的话题。因此，首先要从信息交互的内容学习话题，并得到每个信息交互的话题分布。之后依次执行识别方法的两个阶段。

在第 1 阶段，对信息交互网络进行化简得到有效信息交互网络，具体方法如图 7-2 所示。该方法枚举所有的信息交互，并根据时效和话题相似度检查它们的有效性。这里使用信息交互的内容的话题分布来表示信息交互与不同话题的相似度。明显地，化简过程可以在多项式时间内完成。

一旦得到有效信息交互网络，策略博弈就开始执行。当一个个体被选择进行操作时，他有 4 种可选的操作：加入、退出、变更和不做行动。每种行动具体描述如下。

（1）加入：个体 a_i 加入一个新的信息交互群体，这时，在他使用的策略 s_i 中加入这个新的信息交互群体的话题标识。

Algorithm 7.1　Reduction of the original interaction network.

```
procedure REIN(𝓘,Δ_Sim,Δt)
    for all e∈E do
        if ΔT(e)≤Δt then
            td[]=Topic_Distri(e,Γ(e))
            td_max=Max(td[])
            if td_max≤Δ_Sim then
                Delete e
            end if
        else
            Delete e
        end if
    end for
    for all v∈V do
        if Degree(v)==0 then
            Delete v
        end if
    end for
    𝓘'=𝓘
end procedure
```

图 7-2　交互网络约简算法

（2）退出：个体 a_i 退出一个已加入的信息交互群体的讨论，这时，从他使用的策略 s_i 中移除这个信息交互群体的话题标识。

（3）变更：个体 a_i 从一个已加入的信息交互群体中换到一个未加入的信息交互群体中，这时，从使用的策略 s_i 中移除已加入的这个信息交互群体的标识，加入新换入的这个信息进行交互群体的话题标识。

（4）不做行动：个体 a_i 保持他的策略 s_i 不变。

由于局部均衡总是存在的，当达到局部均衡时，所有的个体都将不会再做任何动作。局部均衡的计算算法如图 7-3 所示。在该算法中，所有个体在初始时都没有加入任何信息交互群体，即他们的策略是空集。每个个体均有一个固定的策略空间，这个策略空间由个体所参与过的话题组成。在每个时步，被选择采取行动的个体会按照加入、退出、变更和不做行动这 4 种动作来评估其策略空间中的策略，并选择最佳的应对方案。

设在某一时步，个体a_i被选择采取行动，$degree(i)$是他的度，$group(s_i)$是其所属的信息交互群体。计算$u_i(s_i)$的时间复杂度依赖于$g_i(s_i)$的时间复杂度。根据a_i的度，g_i的第 1 项可以在$O(degree(i))$时间内完成计算。g_i的第 2 项 $\sum_{j\in n}\frac{d_j}{2m}|s_i\cap s_j|$ 可以变形为 $\sum_{s\in s_i}\sum_{j\in group(s)}\frac{d_j}{2m}$。由于只有个体$a_i$加入或退出$group(s)$，当在每一时步内记录并更新 $\sum_{j\in group(s)}\frac{d_j}{2m}$ 时，就可以保证这一项在固定的时间内计算完成。因此，g_i的第 2 项也可以在固定的时间内完成计算。这样，一个策略可以在$O(degree(i))$的时间内完成评估。另外，对于加入操作，a_i只考虑Z_i-s_i中的话题；对于退出操作，a_i只考虑s_i中的话题；而对于变更操作，a_i则考虑从s_i移除一个话题并从Z_i-s_i中添加一个话题。最终，总的时间复杂度是$O(|s_i|\cdot|Z_i-s_i|\cdot degree(i))$。

Algorithm 7.2 Compute local equilibria.

```
procedure CLE(I')
    for all v∈V' do
        Initialize v's strategy space according to its observed topics z
        Initialize v's strategy s=∅
    end for
    repeat
        a_i=Random_Select(V')
        u_i=Calc_Utility(a_i)
        u_i'=MinValue
        op_i=Null
        for all op∈{join,quit,switch} do
            u_op=Calc_Utility(a_i,op)
            if u_i'<u_op then
                u_i'=u_op
                op_i=op
            end if
        end for
        if u_i<u_i' then
            Update s_i with op
        else
            a_i does nothing
        end if
    until No agent can benefit itself by changing its strategies
end procedure
```

图 7-3　计算局部均衡

7.3.4　实验结果分析

通过将本文的方法与已有的两个基准方法在从 Facebook 上采集的真实数据集上进行实验，以验证本文方法的有效性。评价指标采用适合评估网络中存在重叠现象的扩展的模块化度[36]。

实验中用到的 Facebook 数据集包含自 2011 年 11 月～2013 年 12 月的 11 906 条消息以及这些消息的评论。这些消息通常都比较短，他们的评论都是对原始消息感兴趣的用户发布的。一条消息的发布者和它的一个接受者之间的相互回复与评论成为一个信息交互。最终，构成的信息交互网络含有 7417 个节点和 70 703 条边。

本文方法的两个阶段均依赖于从信息交互内容产生的话题。因此，在实验中使用基于 LDA 话题模型的话题产生工具 gensim 预先生成话题。设 k 是生成的话题总数，Δ_{Sim} 是话题相似度阈值。k 和 Δ_{Sim} 对于本文方法在识别信息交互群体的性能上有着显著的影响。不同的话题数和话题相似度阈值会导致从同一信息交互网络生成不同的有效交互网络。为了降低话题相似度和话题数过大而导致的有效信息交互网络的过于稀疏，通过设置足够大的 Δ_t 来纳入更多的信息交互。这样导致的一个不足是会降低群体的时效性。

信息交互群体的识别关注的是找出那些在相同话题下交互紧密的个体，而并不关心这些个体讨论的具体内容。话题数的设置范围是 4、6、8、10、13、16、20 和 3，而Δ_{Sim}设置的范围为 0.6～0.95，步长是 0.05。

为了验证本文方法的有效性，将其与两种不同的方法进行对比实验。两种对比的方法是社区形成博弈方法(community formation game，简称 CFG)[28] 和 LDA-Word 方法[38]。

(1) CFG 方法是一种基于结构的方法，它在划分网络结构时只使用了结构信息。该方法是一种基于博弈论的方法，也是本文方法的灵感来源。

(2) LDA-Word 方法是一种典型的基于内容的方法，它只是依据节点信息交互的内容的相似性对节点进行划分，把交互内容相似的节点划分为同一社区。

3 种方法均使用相同的信息交互网络作为输入来识别信息交互群体。不同于 CFG 和 LDA-Word 两种方法，本文方法的目标是识别出具有时效的、动态的信息交互群体，因此会有检验信息交互有效性的化简阶段。也就是说，尽管对 3 种方法来说，初始输入的信息交互网络是相同的，但是本文方法在实际进行识别时使用的网络比两种对比方法要小。为了能够使 3 种方法的识别结果在同一尺度下进行比较，把 3 种方法的识别结果在有效交互网络上进行了对比。值得注意的是，CFG 方法和本文方法都具有随机性，因此这两种方法的结果均是它们各自运行 10 次的平均结果，目的是降低随机性的影响。

图 7-4 显示的是 3 种方法在 Facebook 数据集上的对比结果。图中各子图的曲线表示

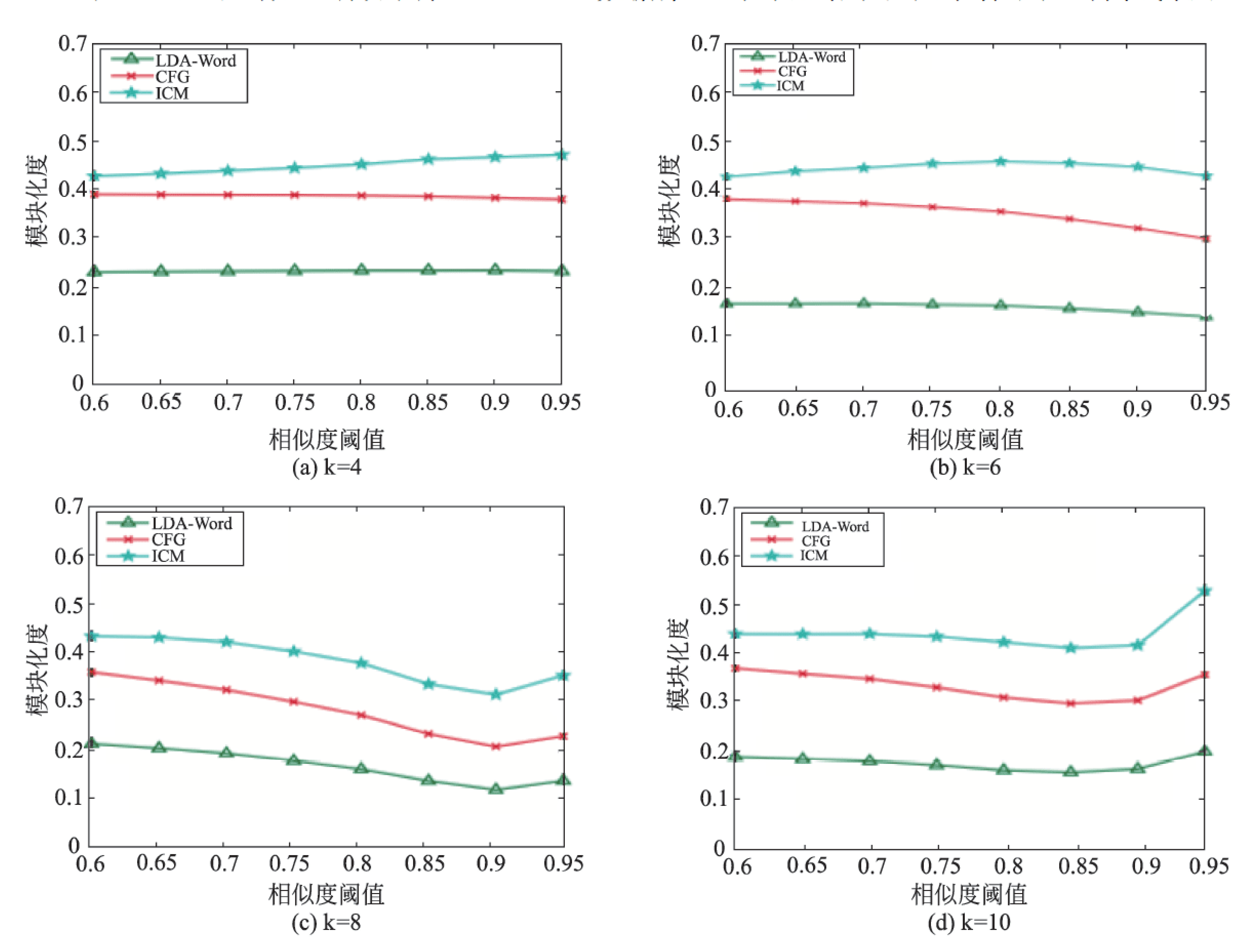


图 7-4 Facebook 数据集上不同方法在模块化度上的表现，ICM 表示的是本文方法

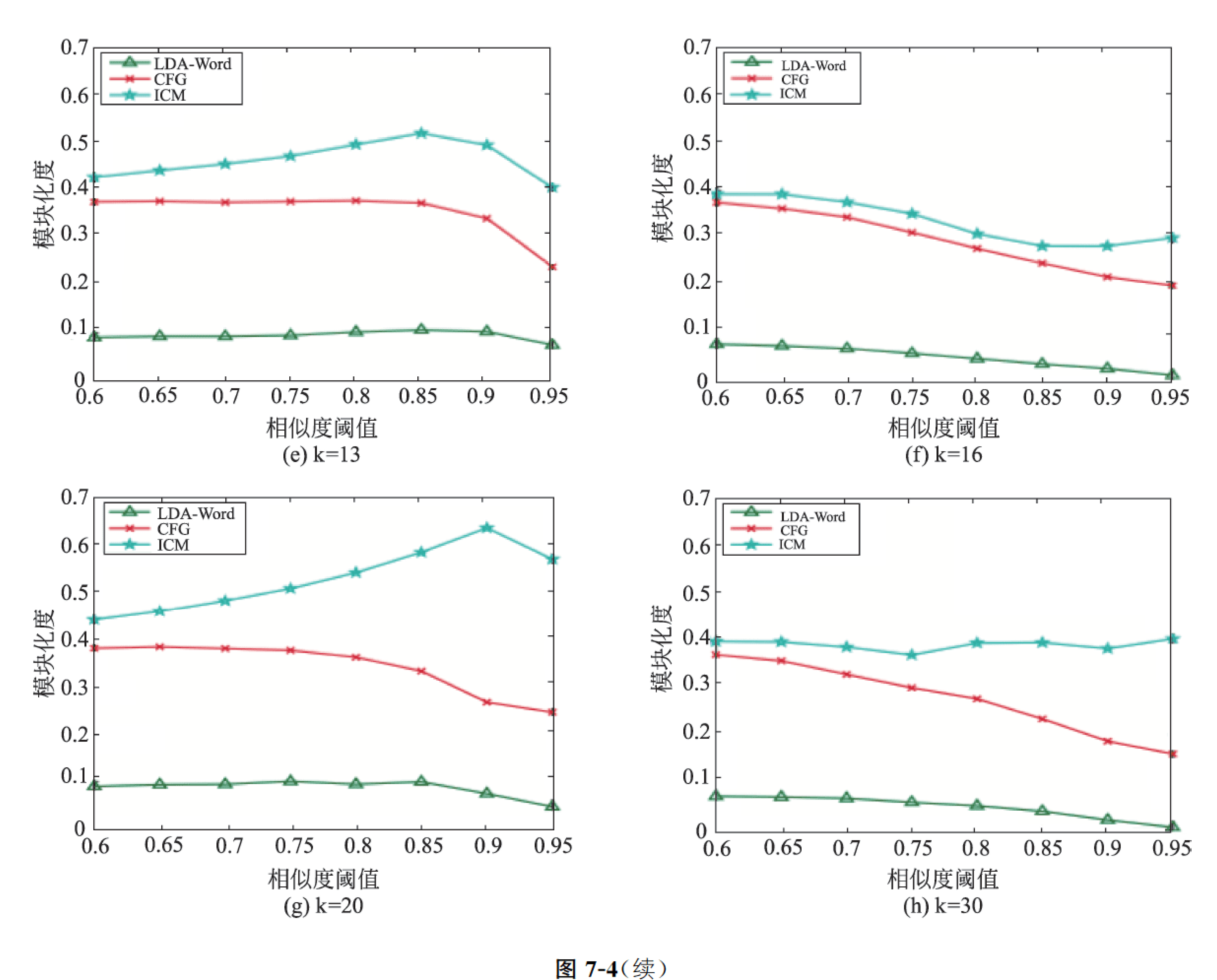


图 7-4(续)

的是在不同的话题数 k 和不同的话题相似度阈值下的模块化度。从图中可以看到,本文方法在每个话题数 k 下的表现都比 CFG 和 LDA-Word 方法要好。从图 7-4(a)～图 7-4(e)可以看到,每个子图中的 3 条曲线都有相似的趋势。而在图 7-4(f)～图 7-4(h)中,本文方法的结果曲线的表现与两种对比方法不同。在这 3 个子图中,CFG 和 LDA-Word 方法的模块化度都随着话题相似度阈值的变大而减小。相比而言,即使在话题相似度的阈值很大的情况下本文方法的模块化度并未减小。实验的结果显示,本文的方法即使在话题数和话题相似度阈值很大的情形下仍能找到交互紧密的群体,具有较好的表现。

7.4　重叠群体的行为与结构协同演化

从前面几章对社交网络中个体信息行为的演化分析可以看到,个体的行为(包括个体信息行为和关系调整行为等)会促使社交网络中的个体围绕着信息的内容、个体的偏好等因素产生聚集现象,形成不同的群体。

在现实世界中,人们总是会身处不同的环境中,拥有不同的社会关系网。举例来说,一个人会有由家人亲戚组成的亲属关系群体,有由同学组成的同学关系群体,有由朋友组成的好友关系群体,有由同事、客户等组成的工作关系群体等。同样地,在社交网络中个体会根

据自己的兴趣、爱好等因素成为不同群体的成员，而这些不同的群体由于共同成员的存在成为重叠群体。重叠群体中的共同个体使得不同的群体之间产生了联系。这些群体通过共同个体相互间产生影响和作用。例如，在现实世界中，人们从好友关系群体中获得的信息可能会影响到他们在工作关系群体中的行为与决策。类似地，在社交网络中，个体从一个群体中获得信息也会影响到他在其他群体中的行为。本节将初步探讨社交网络中存在相互影响、相互作用的重叠群体（多群体）的演化性。

7.4.1　重叠群体相互影响的方式

为了能够简单、明了地说明重叠社交群体的相互影响，不失一般性，本节使用两个重叠的群体进行说明。

图 7-5 显示了两个群体以及它们之间的重叠关系，这两个群体分别表示为群体 A 和群体 B，其中的绿色个体表示它们的相互重叠的个体。

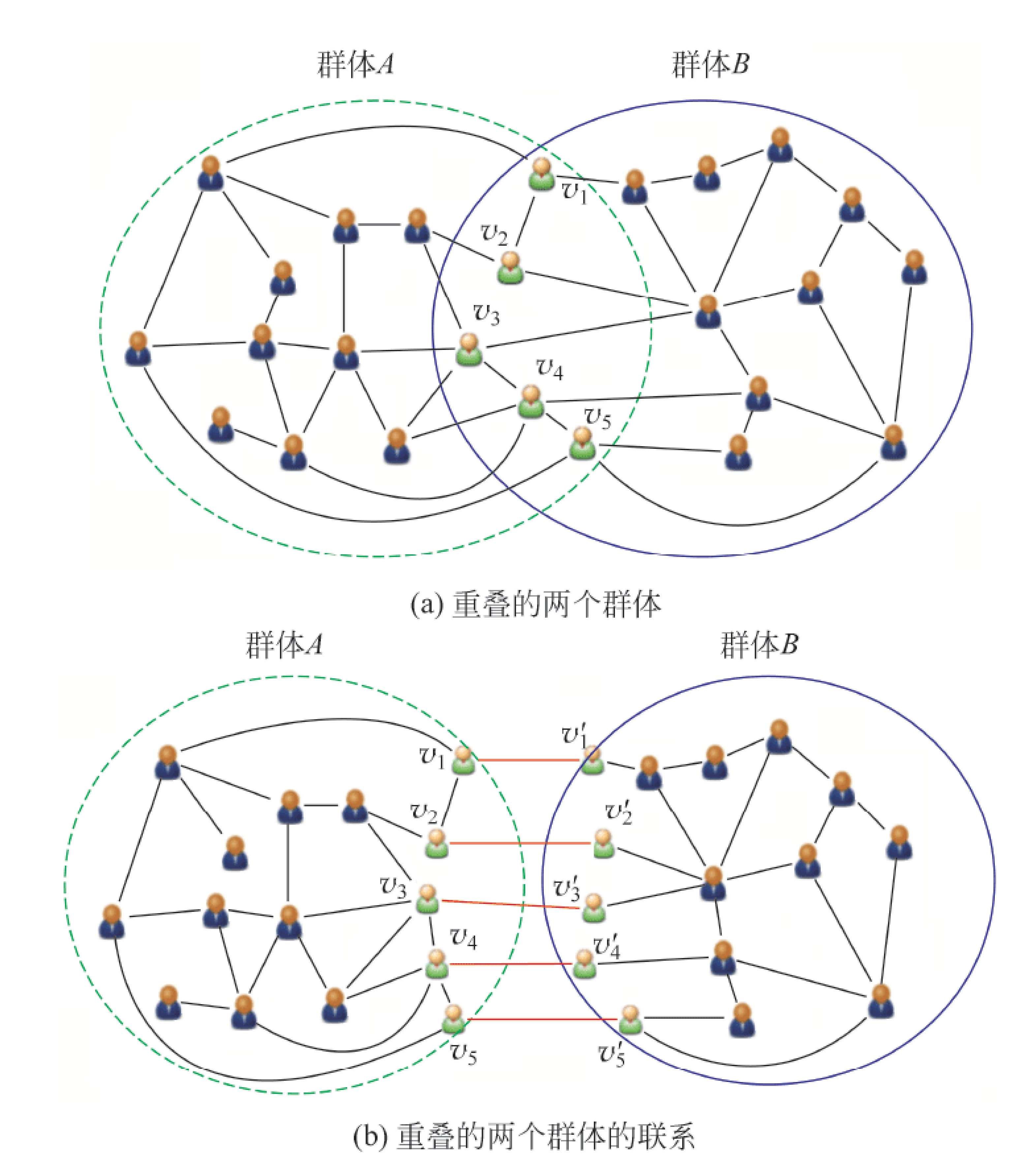


图 7-5　重叠群体的联系，两个群体之间的红色连边表示了由共同个体所产生的相互联系

图 7-5(a)说明了群体 A 和群体 B 的重叠关系。在该子图中，标识为$v_1 \sim v_5$的 5 个绿色个体是两个群体的重叠个体，它们既和群体 A 中的个体有社交关系，也和群体 B 中的个体有社交关系。

由于群体 A 和群体 B 是不同的两个群体，这导致了重叠个体在群体 A 和群体 B 中的个体标识（如角色等）也不相同。因此，图 7-5(a)所示的重叠关系可以进一步地表示成图 7-5(b)所示的关系。在图 7-5(b)中，重叠个体在群体 A 中表示为 $v_1 \sim v_5$ 的绿色个体，而在群体 B

中表示为v_1'～v_5'的绿色个体。

以v_1所表示的个体为例，该个体在群体 A 中的个体标识为v_1，其在群体 B 中对应的个体标识为v_1'，v_1和v_1'之间的红色连边则表示了其共同个体的身份。对于两个群体来说，这些红色连边则表示了它们之间由共同个体产生的相互联系。

相互影响的群体的影响形式可以分为两类：单个体影响和群体影响，如图 7-6 所示。需要注意的是，这里，群体 A 和群体 B 是对等的，无主从之分。该图以群体 A 对群体 B 的影响进行说明，反过来，群体 B 对群体 A 的影响也是一样的。单个体影响如图 7-6(a)所示，当群体 A 中的个体v_1与个体v_2进行相互作用时，如博弈、策略更新、关系更新等，v_1只受到其在群体 B 中的对应个体v_1'的影响。与单个体影响不同，群体影响时，群体 A 中的个体v_1受到群体 B 中以对应个体v_1'为中心的群体的影响，即受到v_1'及v_1'的邻居个体的影响，如图 7-6(b)所示。

下一节将讨论在上述存在相互影响的多群体中的社交演化博弈模型。

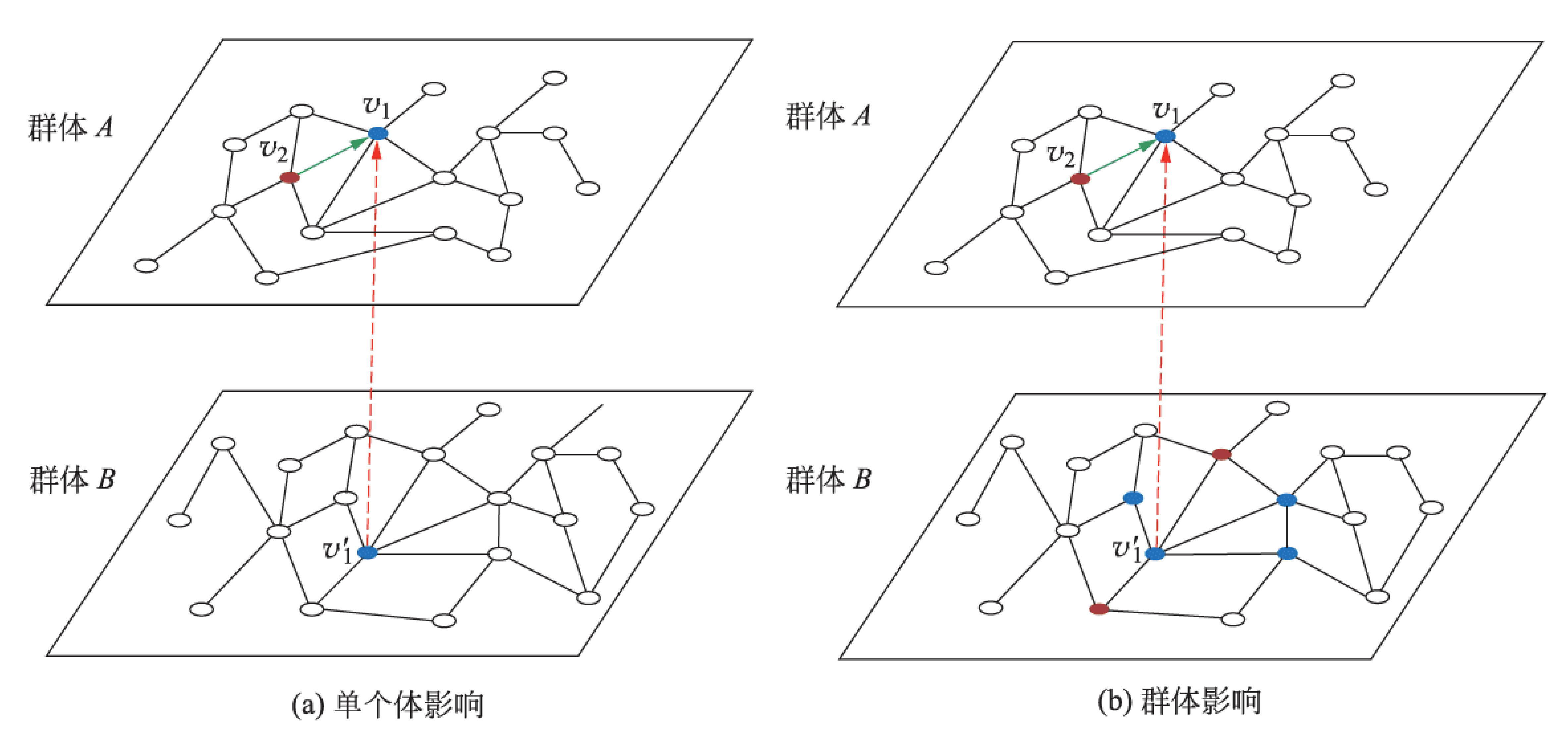


图 7-6　相互影响群体的影响形式

蓝色个体表示合作者，红色个体表示背叛者。红色箭头虚线表示了群体间的影响。群体 A 和群体 B 是对等的，因此两个群体的影响关系反过来也是成立的

7.4.2　模型设置

不失一般性，本章使用采用无权无向图来表示群体。相互影响的群体 A 和 B 均是大小为 N、节点平均度数为 k 的无权无向的随机网络。群体 A 与群体 B 之间的相互影响是对等的，即群体 A 对群体 B 的影响与群体 B 对群体 A 的影响是对等的，没有主从之分。群体 A 和群体 B 之间的相互影响是部分影响的，设相互影响的节点的比例为 ρ，即群体 A 和群体 B 中均有 $N'=\rho N$ 个相互影响节点。群体 A 中的节点记为 i，$i\in N$，群体 B 中的节点记为 i'，$i'\in N$。两个群体中节点均附加判断是否是相互影响节点的属性 Idp。如果 $Idp(i)=1$ 或 $Idp(i')=1$，则群体 A 中的 i 与群体 B 中的 i'相互影响；反之，如果 $Idp(i)=0$ 或 $Idp(i')=0$，则群体 A 中的 i 与群体 B 中的 i'不相互影响。

两个群体上的策略博弈均设为囚徒困境。初始时，群体 A 和群体 B 上的节点都被等概率地分配角色，要么是合作者(C)，要么是非合作者(D)。两个群体上节点的效用π_i和$\pi_{i'}$的

计算方法是一样的。下面以群体 A 为例,对其上进行说明(对群体 B 也适用)。

群体 A 中的每个节点均与其最近的邻居节点进行囚徒困境博弈。设节点 i 与其最近的邻居节点进行博弈,当两个节点都是合作的,则他们的收益都是 $R=1$;当两个节点都是非合作的,则他们的收益都是 $P=0$;当一个节点是合作,而另一个节点是非合作的,则合作者的收益是 $S=0$,非合作者的收益为 $T>1$。收益矩阵,如下所示:

$$M=\begin{pmatrix}1 & 0\\ T & 0\end{pmatrix}$$

则 i 的效用计算方法如下:

$$u_i=\sum_{j\in\mathcal{N}_i} s_i^{\mathrm{T}} M s_j$$

其中,$\mathcal{N}_i$是 i 的邻居节点集合,s_i 和 s_j 分别是 i 和 j 使用的策略,M 是公式所示的博弈收益矩阵。

群体中的每个节点均有声誉,它反映了节点在社交演化博弈过程中的合作情况。

$$R_i(t)=R_i(t-1)+\Delta_i(t)$$

其中,当 i 在 t 时刻是合作的情形下,$\Delta_i(t)=1$,否则,$\Delta_i(t)=0$。

社交演化的更新机制是异步的,在演化过程中,每个节点都有机会进行策略更新和社交关系更新。在每个演化周期 t,群体中随机的一个节点进入更新状态,即策略更新或社交关系更新。

设节点 i 进入更新状态,以概率 W 进行策略更新,$W=\tau_s/\tau_e$,τ_s是策略更新的时长,τ_e是社交关系更新时长。社交关系更新的频率随 W 的增大而增大。

(1) 策略更新

群体 A 和群体 B 之间不存在实际的物理连接,它们之间的相互影响是通过存在影响关系的节点在策略更新的影响上体现的。节点 i 进行策略更新时,他从其邻居节点中选择效用最大的一个节点 j,并按下式确定的概率模仿 j 的策略。

$$W(s_i\leftarrow s_j)=w_i\,\frac{1}{1+\exp[\beta(\pi_i-\pi_j)]}$$

其中,w_i是群体 B 中 i'对群体 A 中 i 的策略更新的影响。

当相互影响群体的影响是单个体影响时,w_i按下式确定:

$$w_i=\begin{cases}w_{\max}, & s_i=s_{i'}\\ w_{\min}, & s_i\neq s_{i'}\end{cases}$$

当相互影响群体的影响是群体影响时,w_i则按下式确定:

$$w_i=1-(w_{\max}-w_{\min})\frac{N^{s_i}}{G}$$

其中,G 是群体 B 中对 i 有影响的节点集(由 i'和 i'的邻居节点组成),N^{s_i} 是 G 中与 i 策略一样的节点个数,这里,w_i是支持 i 维持自身策略的程度。

(2) 网络结构更新

群体 A 中的节点之间的社交关系会随着时间而发生变化。当群体 A 中的一个节点进行社交关系更新时,他寻找一个未建立关系的节点建立社交关系,同时切断一个已有的社交关系。具体地,当节点 i 建立新的社交关系时,他从其邻居节点的邻居节点中,依概率p_r按照声誉最大原则选择一个新节点 j 作为其邻居节点建立社交关系。

$$k = \arg \max_{\mathcal{N}_i^I \cup_{l \in \mathcal{N}_i^O} \mathcal{N}_l^O \setminus (\{i\} \cup \mathcal{N}_i^O)} \{R_l(t) > R_i(t)\}$$

或者以概率 $1-p_r$ 随机选择一个新节点 j 建立社交关系。当节点 i 切断一个已有的社交关系时，他断开与邻居节点中声誉最低的节点的社交关系。

$$j = \arg \min_{l \in \mathcal{N}_i^O} \{R_l(t) < R_i(t)\}$$

7.4.3 实验结果与分析

初始时，两个相互影响的群体是大小均为 $N=2000$ 的随机网络，平均 $k=8$。使用合作者比例 f_c 随时间的变化来分析两个群体的相互影响对群体演化的影响。演化仿真的时长为 2×10^5 个周期，此时认为演化达到相对稳定状态。为了减小随机性的影响，每组数据都是独立仿真 10 次后计算的平均结果。

首先，分别在单个体影响和群体影响的条件下观察 f_c 在不同的 W 和 ρ 下的演化。图 7-7(a)显示的是在单个体影响的情况下，相互影响的群体 A 和群体 B 组成的系统中 f_c 随时间的变化曲线。从该图中可以看到，所有的 f_c 演化曲线都呈现出“先下后上”的形状，这显示出系统中存在着网络互惠的情况，并发挥了促进合作的作用。在同一 W 下，$\rho=1$ 的 f_c 曲线的转折点比 $\rho=0$ 的 f_c 曲线的转折点更低，出现的时间更晚。这说明，在相互影响群体中，合作个体需要更多的时间进行调整和聚集。在这一时期，合作者要承受背叛者更多的入侵。$\rho=0$ 时，演化曲线的转折点的 f_c 值随着 W 的增大而变大。同样地，在 $\rho=1$ 时，演化曲线的转折点的 f_c 值也随 W 的增大而变大。群体中个体调整关系的频率变大可以帮助合作者抵御背叛者的入侵。图 7-7(b)显示的则是在群体影响的情况下，系统中 f_c 随时间的变化曲线，可以看到，群体影响的情形与单个体影响的情形类似，合作个体也需要更多的时间进行调整和聚集。需要注意的是，在图 7-7 所示的参数设置中，当网络中不进行个体关系调整($W=0$)时，系统达到相对稳定状态时合作个体与背叛个体并存。

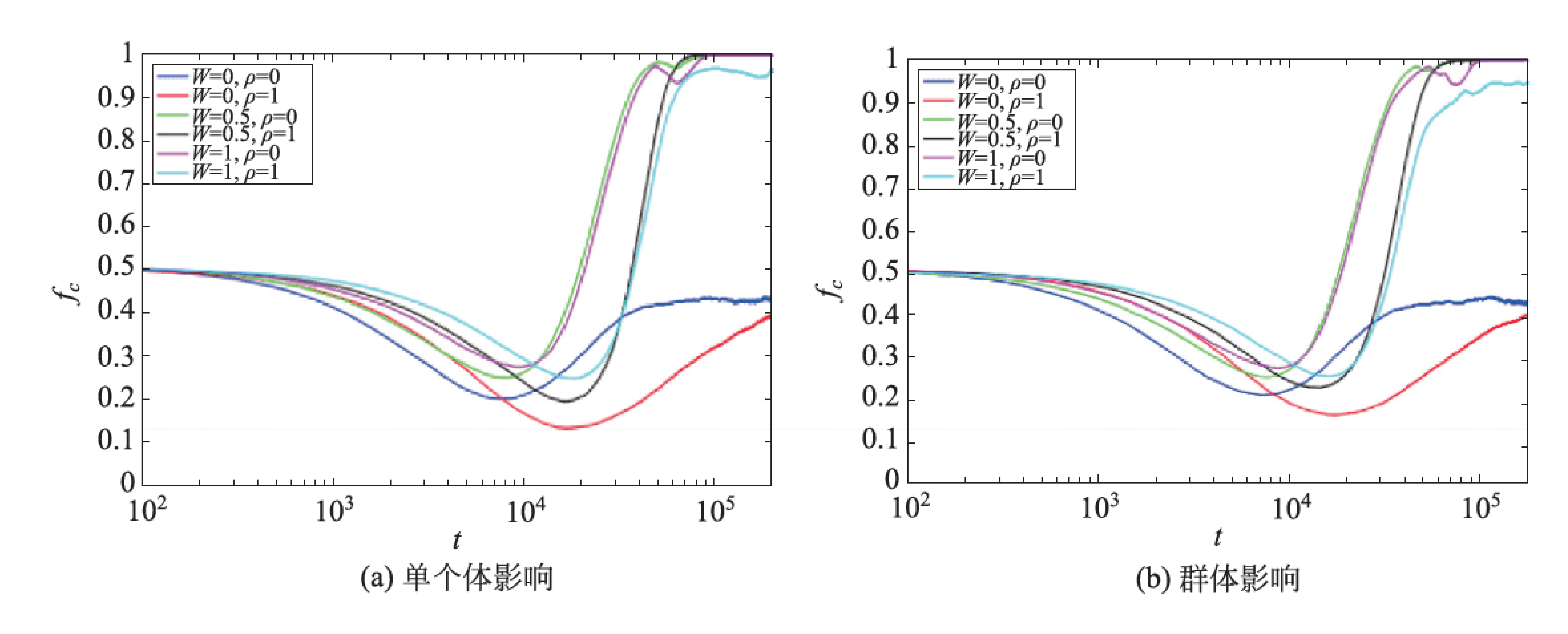


图 7-7 单个体影响和群体影响时 f_c 随时间的演化

$N=2000$，$\langle k\rangle=8$，$b=1.6$，$\beta=1$，$p_r=0.5$，$p_s=0.1$，$\sigma=1$

图 7-8 的 3 个子图分别显示了不同 W 时单个体影响与群体影响情况的 f_c 值随时间变化的曲线对比情况。通过对比图 7-8(a)～图 7-8(c)可以观察到，$\rho=0$ 时的单个体影响曲线和群体影响曲线是高度重合的。这是因为，$\rho=0$ 时群体 A 和群体 B 之间没有影响，单个体

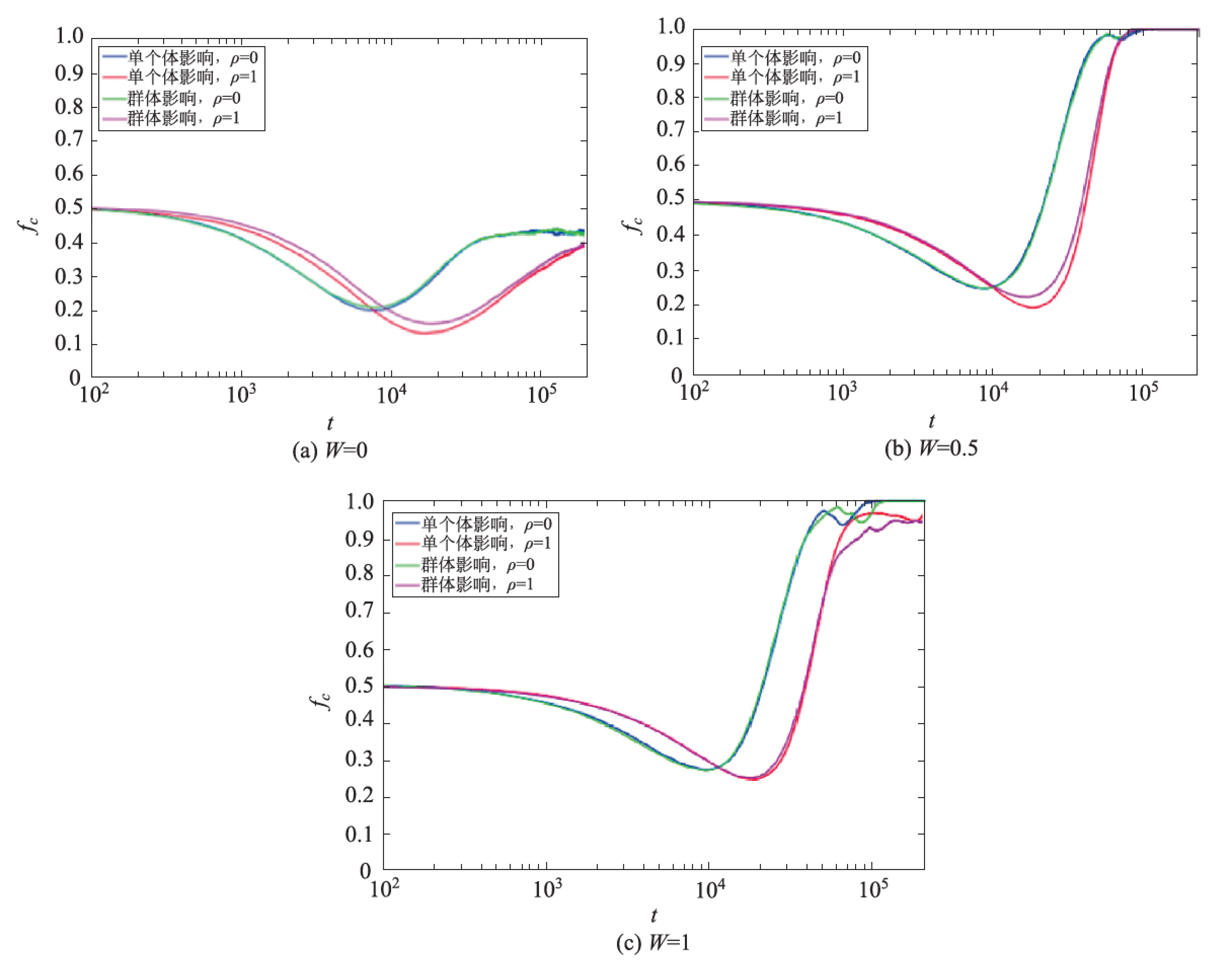


图 7-8　不同 W 的条件下f_c随时间的演化

$N=2000$，$\langle k\rangle=8$，$b=1.6$，$\beta=1$，$p_r=0.5$，$p_s=0.1$，$\sigma=1$

影响和群体影响的设置不起任何作用，社交演化博弈是群体 A 和群体 B 中各自独立进行的。同时也可以观察到，当个体策略与关系都发生演化时，单个体影响与群体影响对合作演化的影响的差别比它们在方格网络上对合作演化的影响的差别要小。

7.5　跨网络群体的用户关注度竞争

过去 10 年间，以微博、社交网站、微信为代表的社交网络服务一直在以极高的速度扩张，其扩张速度多年间一直维持在 100%左右[39]，每年都会有数以亿计的新用户涌入社交网站中。更重要的是，有大量已经在使用社交网络服务的用户投入了更多的时间和精力到已经参与的社交网络服务，或者更新型、更有趣的新的社交网络服务中。该发展阶段可命名为发展中阶段。

然而，当社交网络中的用户数量迫近世界网民总量的时候，获取新用户或者榨取老用户的更多使用时间变得越来越困难，那么可将该发展阶段命名为社交网络服务的发达阶段。当前在线社交网络已经进入了发达阶段，新的用户资源变得越来越稀缺，社交网络运营商被迫要与其他社交网络运营商竞争，从其他运营商处获取更多的用户关注度，从而保证自己在行业中的地位。

许多前人关于社交网络演化的研究成果集中于描述社交网络服务用户及相关的网络结构。由 Yu 等人[40,41]提出的社交网络演化博弈(SEG)模型主要考虑两大因素：用户的短期收益和长期名声。用户的短期收益控制用户的信息交互行为，通常可以用合作博弈模型来描述。如果一个用户选择接受并转发一条邻居的消息，该行为会被认为是一种合作行为；而如果一个用户不去进行转发，则会被认为是一种背叛行为。若两个用户均选择合作，则两个用户均获得收益 R；若两个用户均选择背叛，则其收益均为 P。一个背叛者可以从合作者处获取收益 T，而被背叛者的收益为 S。通常来说，设定 $T>P>R>S$，并用用户的长期名声控制社交网络的结构。如果一个用户总是不能与其他用户合作，则在一定情况下，其邻居用户可以有机会断绝朋友关系。还可以通过引入更多的属性使得本模型更加实用，比如引入混合策略来描述信息交互博弈过程等[42]。

然而，描述社交网络本身的行为和策略的工作还相对较少，而这对于在发达阶段的描述其实具有非常重要的意义。Wu 等人[43]研究了社交网络和社交网络用户之间关于用户隐私相关事项的演化博弈模型，虽然对用信息传播策略来说社交网络服务商不能直接干涉用户的行为，但他们可以改变服务规则，例如通过增加更多的合作策略收益来鼓励更多的合作策略。另外，网络结构的初始状态同样可以影响社交网络的演化结果。然而，描述在同一个用户集合下的两个或更多个社交网络之间的演化博弈的工作仍然不足，特别是当用户的数量稳定且用户花费在社交网络上的时间基本不变的时候。

本章提出了一种竞争性社交演化博弈模型(competitive social evolutionary game，简称 CSEG)来描述在一个用户集合之上的多个社交网络服务竞争过程，如图 7-9 所示。通过 CSEG 可以仿真不同的社交网络初始状态与信息传播规则对社交演化过程的影响，以及它们如何影响到每个用户在各个社交网络上时间的分配情况。这里假设一个网民花费在社交网络服务上的时间是一个固定值，因此，如果其在某个社交网络上花费更多的时间，则其在其他社交网络上所花费的时间就会变少。这些用户在社交网络上的行为主要包括转发邻居的消息以及增加或者删除邻居。在此引入用户关注度的概念：一个社交网络用户在不同社

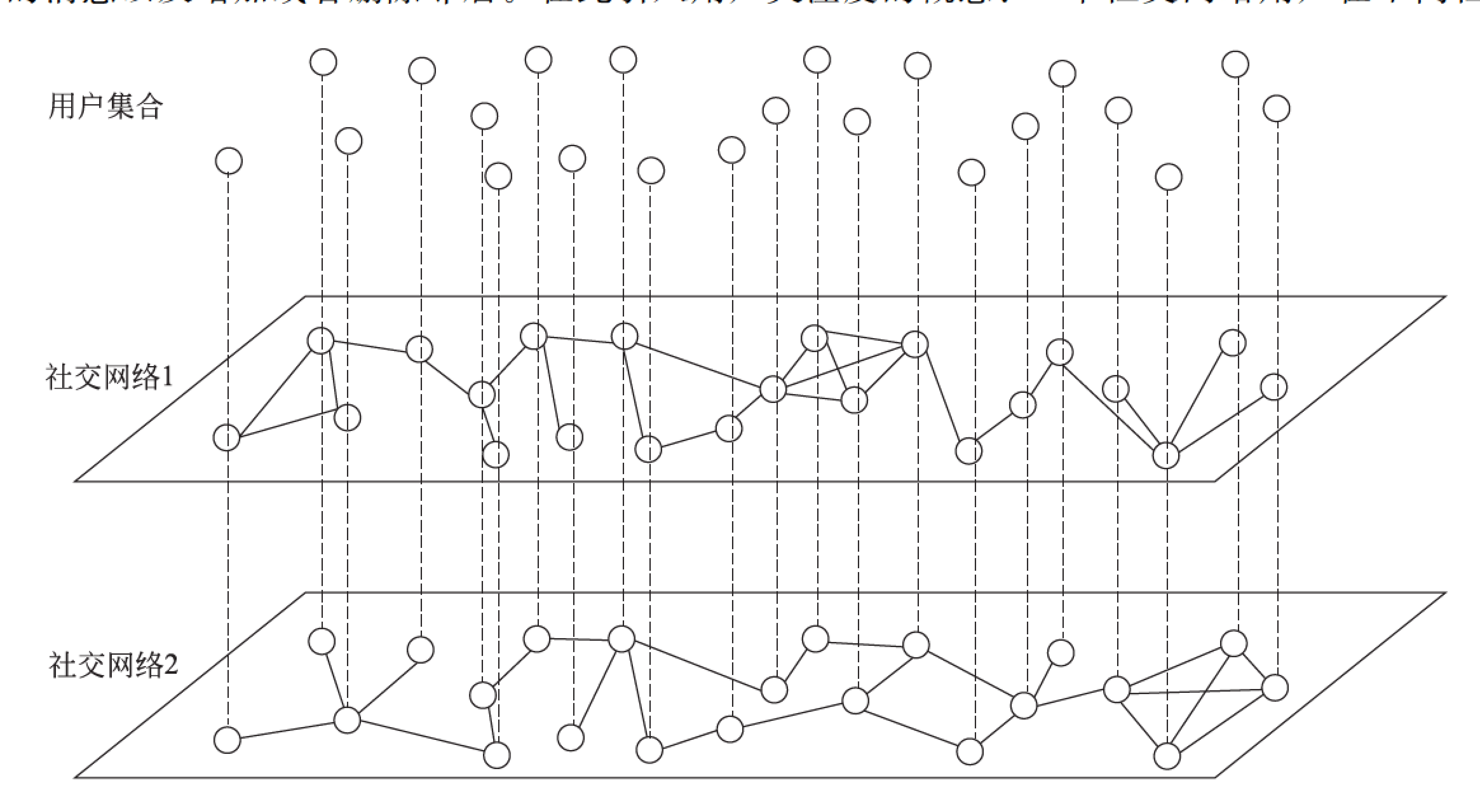


图 7-9　基于同一组用户的两个社交网络示意图

交网络上的访问时间的分配比例。假定社交网络服务商的竞争本质上来讲就是对用户关注度的竞争,则研究目标就被确定为分析在不同初始条件和策略下的用户关注度,包括不同的初始网络结构、合作博弈的收益矩阵以及用户的邻居更新策略等等。

7.5.1 竞争性社交演化博弈模型

本节介绍描述单一用户群体上的多个社交网络间的竞争性社交演化博弈模型(CSEG)。在单个社交网络中,假定每个用户关注的内容包括两个方面:短期的收益(utility)和长期的声望(reputation)。基于这两个关注的内容,提出了两类更新机制:为获得更好收益的策略更新机制以及为获得更好声望的邻居更新机制。而用户集合的关注度情况的描述则由关注度矩阵 A 来描述。

7.5.1.1 博弈模型

一个竞争性社交演化博弈模型由六元组 $G=(V,Z,I,U,R,A)$组成:

- $V=\{v_1,v_2,\cdots,v_n\}$为用户的集合;
- $Z=\{E_1,E_2,\cdots,E_m\}$为 m 个社交网络的边集合,其中,$E_k=\{e_{ij}\mid v_i,v_j\in V\}$是用户集在社交网络 SN_k 上的朋友关系集合,$k=1,2,\cdots,m$;
- I 为用户根据其策略所执行的博弈行为,包括提高短期收益与长期声望的行为;
- U 为用户短期收益函数的集合;
- R 为用户长期声望函数的集合;
- $A_{mn}=\{a_1,a_2,\cdots,a_n\}$为 n 个用户在 m 个社交网络上的用户关注度矩阵。

(1) 有向图和无向图

如果一个社交网络的用户间关系为有向关系,如微博的关注关系,那么一个用户 i 有两类邻居:入度邻居和出度邻居。若存在边 e_{ji},则用户 j 是 i 的入度邻居,并可用 $i\leftarrow j$ 表示。相对应地,i 是 j 的出度邻居。令 N_i^I 表示 i 的入度邻居集合,N_i^O 表示 i 的出度邻居集合,则 i 的邻居集合可以表示为 $N_i=N_i^I\cup N_i^O$. d_i^I 和 d_i^O 为 i 的入度和出度,$d_i=d_i^I+d_i^O$ 为 i 的度。

如果一个社交网络的用户间关系为无向关系,如 Facebook 等社交网站的朋友关系,则该社交网络只有邻居 N_i 和度 d_i 的概念。

(2) 用户关注度

在本节概述中给出了用户关注度的概念,其涵盖了一个用户在所有其参与的社交网络上的参与时间的分配情况,对于一个用户来说,其在特定时刻只会有部分精力关注于某一个社交网络。且由于本文假设社交网络服务已经进入了发达阶段,用户的数量和用户在社交网络上所花费的时间均相对稳定。

为更加形式化地描述用户关注度,定义**用户关注度矩阵如下**:

$$A_{mn}=\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

A_{mn}代表一个 n 个用户 m 个社交网络的用户关注度,$a_{ij}\in A_{mn}$为用户 j 在社交网络 SN_i

上的关注度。对于所有用户 $j\in n$，矩阵中的每一列之和 $\sum_{i=1}^{m} a_{ij}=1$，分别代表每个用户在所有社交网络上花费的全部时间。

此外，定义 SN_i 的流行度 a_i^* 来刻画 SN_i 该社交网络的成功程度(市场占有程度)，其计算方法为用户关注度的平均值 $a_i^*=1/n\cdot\sum_{k=1}^{n} a_{ik}$。进一步定义所有社交网络的成功程度矩阵 $p=[a_1^*,a_2^*,\cdots,a_m^*]$。

(3) 社交网络用户效用

对于用户效用的计算，CSEG 与 SEG 采用同样的计算方法，用户 i 的效用是在一轮中与所有邻居/对手博弈的累积收益。用户效用是用户的短期收益，影响用户模仿其他用户策略时的选择结果。博弈矩阵可以是一个囚徒困境博弈，矩阵的值既可以是仅考虑该社交网络服务商的设置，也可以是同时考虑该用户的邻居或者其他社交网络设置的一系列效用函数。

(4) 社交网络用户的声望

在 SEG 中，用户 i 的声望为 i 的邻居对他的评价，该评价是邻居们对 i 的历史行为的累积评价。声望是用户的长期收益，会影响用户在选择新的邻居时的决定。t 时刻用户 i 的声望可以被形式化地表示为

$$R_i(t)=\sigma R_i(t-1)+\Delta R_i(t)$$

其中，$\Delta R_i(t)$是声望在 t 时刻的增量，σ 是声望的遗忘比率。定义

$$\Delta R_i(t)=\sin\left(\frac{\pi}{2}\cdot\frac{n_i^C}{d_i}\right)$$

其中，n_i^C 代表 i 的邻居中选择合作策略的数量，正弦函数被用来抑制声望的增长率。

CSEG 中的声望计算方法与 SEG 的不同在于与所有社交网络均相关的用户关注度，对于用户的声望及增删邻居的策略会产生重要的影响。

7.5.1.2　跨社交网络协同演化动力学

在每一个博弈轮次 t，用户以一种异步的方式依次在每一个社交网络内展开行动，两个社交网络的协同演化博弈过程如图 7-10 所示。每一轮会执行 3 个步骤：策略更新、邻居更新、囚徒困境博弈。两类更新过程按照异步更新方式执行：在时刻 t_{ij}(在第 i 轮社交网络 j 的行动时刻)，一个随机选择的用户执行更新操作，该用户以概率 $1/(1+W)$更新其策略，否则，该用户更新其邻居。W 是策略更新速率 τ_e 和邻居更新策略 τ_p 的比值：$W=\tau_e/\tau_p$。因此，用户邻居更新的频率会随着 W 的增长而增长。

(1) 策略更新

当被选中的用户执行策略更新时，该用户改变其合作/背叛的基本策略。策略更新完全为用户的个人属性，因此不会受到用户关注度的影响。

当用户 i 决定更新策略时，该用户选择一个满足如下条件的用户 j 并模仿其策略：

$$j=\arg\max_{l\in N_i^O}\{u_l>u_i\}$$

模仿策略可以由 Fermi 更新规则描述：

$$\omega(s_i\leftarrow s_j)=\frac{1}{1+\exp[\beta(u_j-u_i)]}$$

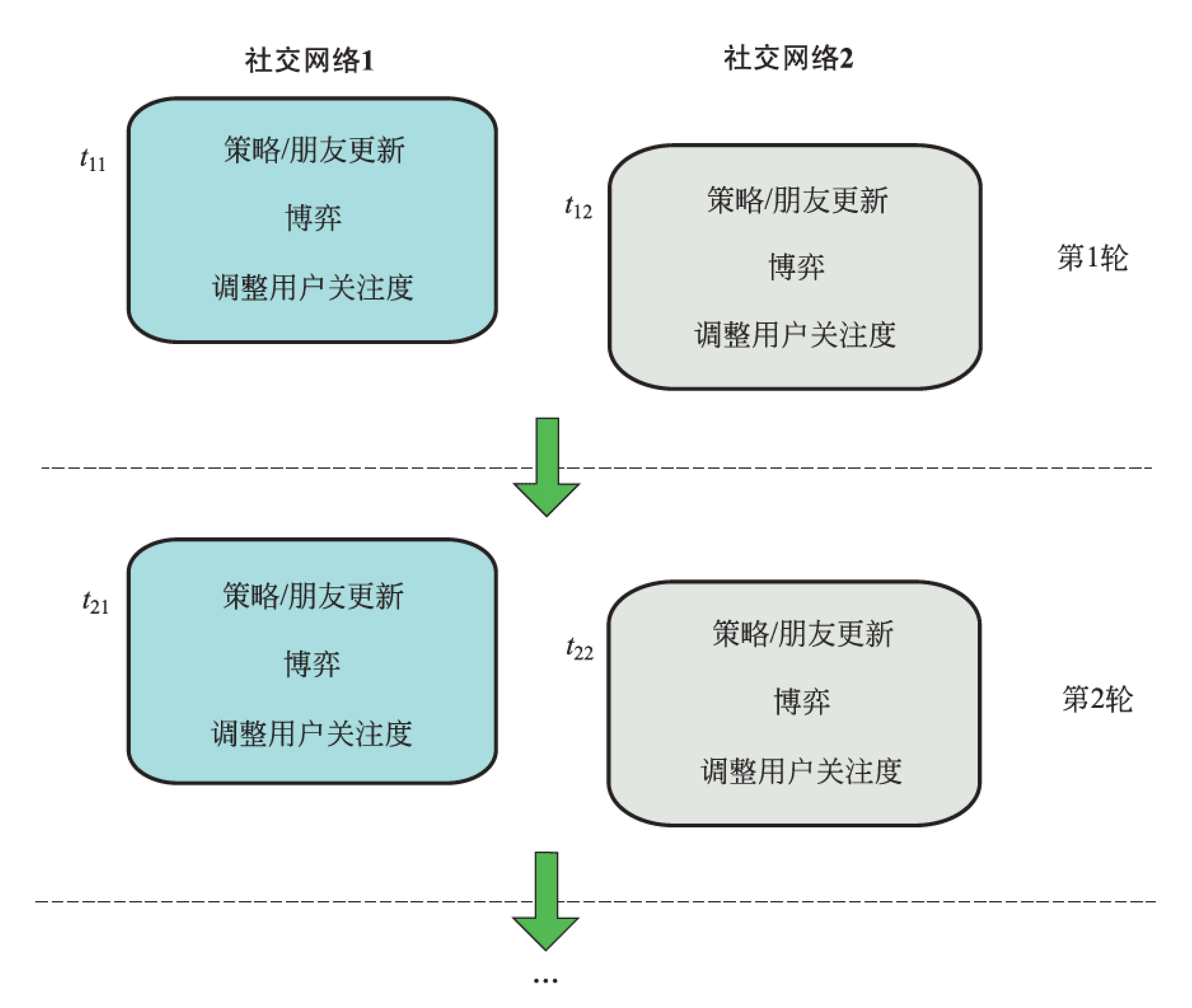


图 7-10　跨社交网络演化动力学

其中，β 代表模仿的噪声，用来描述用户模仿其他用户策略的意愿。

(2) 邻居更新

当被选中的用户执行邻居更新时，该用户在调整邻居时只能意识到其邻居以及邻居的邻居的本地声望信息。在每一轮，用户将声望最低的一个或多个邻居移除，并将其邻居的邻居中声望最高的一个或多个邻居加入其邻居列表中。由于 CSEG 的目标在于研究发达阶段的社交网络服务，因此设定新增和移除邻居的比例应大致为 1。

用户 i 根据最小声望规则移除其一个出度邻居，然后根据最大声望规则增加一个他的入度邻居，或者其出度邻居的邻居。具体地，在概率 p_s 下，用户 i 移除用户 j 满足条件：

$$j = \arg\min_{l \in N_i^O} \{R_l(t) < R_i(t)\}$$

另外，用户 i 在概率 p_r 下，选择满足如下条件的用户 k 作为新的邻居：

$$k = \arg\max_{l \in \{N_i^I \cup \{k \mid k \in N_m^O, \forall m \in N_i^O\}\}} \{R_l(t) > R_i(t)\}$$

否则，在概率 $1-p_r$ 下，用户 i 随机地从入度邻居和出度邻居的邻居中选择用户 k' 作为其新邻居。

(3) 博弈过程

在更新策略之后，对于每一个社交网络，用户 j 与其每一个邻居依次进行囚徒困境博弈，博弈结果同时会受到用户关注度矩阵 A 的影响。j 的效用为其与所有邻居的收益之和，可由下面的公式表示：

$$u_i = \sum_{j \in N_i} s_i^{\mathrm{T}} M s_j$$

其中，N_i 为 i 的邻居集合，s_i 和 s_j 为 i 和 j 所使用的策略，$C=[1,0]^T$ 代表合作策略，$D=[0,1]^T$ 代表背叛策略。矩阵 M 是一个 2×2 的收益矩阵：

$$M=\begin{bmatrix}1 & 0\\ b & 0\end{bmatrix}$$

其中，若两个用户共同合作的话，会获得奖励 1，而背叛一个合作者的收益为 $1\leqslant b\leqslant 2$。

(4) 关注度的调整

在每一演化轮次的最后，全部用户都会调整他们的用户关注度。用户 i 会根据其在不同社交网络上的邻居的合作比例 $f_{ki}(C)=n_{ki}^{C}/d_{ki}$ 来调整该用户的关注向量 $a_i=[a_{1i},a_{2i},\cdots,a_{mi}]^T$。如果

$$f_{ki}(C)\leqslant\frac{1}{m}\sum_{j=1}^{m}f_{ji}(C)$$

那么，a_{ki} 就应该在下一轮被设置为一个更小的值；否则，a_{ki} 应被设置为一个更大的值。用户 i 在各个不同社交网络中的当前效用也会影响到用户关注度。

在计算完用户关注度矩阵 A_{mn} 后，就可以计算用以比较社交网络成功程度的成功度向量 $p=[a_1^*,a_2^*,\cdots,a_m^*]$。

7.5.2 实验结果与分析

本节仿真两个社交网络之间对于用户关注度的竞争博弈过程，其中一个网络初始设置为均匀度网络，用以仿真现实中朋友关系较为均匀的社交网站等服务；另一个网络初始设置为小世界网络，用以仿真现实中朋友关系较为紧密和不均匀的微博等服务。

7.5.2.1 两个社交网络间的用户关注度计算方法

(1) 用户关注度对策略更新的影响

用户个人的效用损益是一个由社交网络服务商确定的社交网络用户的固有属性，因此不会被用户关注度所影响。然而，用户声望却可以被用户关注度影响。例如，当一个用户在社交网络 1 上所花费的时间远比在社交网络 2 上所花费的时间长时，则其在社交网络 1 上的声望重要程度就应相应地比在社交网络 2 上更重要，也就是说，p_r 值应该更高。

在本次仿真中比较两个而非多个社交网络之间的竞争性情况，因此用户 k 的下一轮的声望重视程度指标 p_r' 应按照下面的公式进行计算：

$$p_r'=\begin{cases}p_r+\delta_{\max}(a_{1k}-a_{2k}), & 0\leqslant p_r'\leqslant 1\\ 0, & p_r'<0\\ 1, & p_r'>1\end{cases}$$

其中，$\delta_{\max}$ 为单轮次中声望重视程度的最大变化值。

(2) 用户关注的更新

用户关注的更新与两个因素相关：该轮两个社交网络中的用户声望以及合作/背叛用户的比例。用户 k 在社交网络 1 上的新用户关注度可以被表示为

$$a_{1k}'=\alpha a_{1k}+\beta\frac{u_{1k}}{u_{1k}+u_{2k}}+\gamma\frac{f_{1k}(C)}{f_{1k}(C)+f_{2k}(C)}$$

其中，α,β,γ 为重要程度系数，$\alpha+\beta+\gamma=1$。上式可以保证用户 k 的两项关注度 $a_{1k}+a_{2k}=1$。

7.5.2.2　仿真设置与假设

本仿真过程中的节点数量为 500 个，初始的用户关注度矩阵可以按照下式表示：

$$A_{2\times 500}=\{a_{ij}=0.5\,|\,i\in\{1,2\},j\in\{1,2,\cdots,500\}\}$$

使用 igraph 构建实验中所需要的图，其中小世界网络由 Watts-Strogatz 模型产生，网络中节点的平均度为 4。演化博弈的最大轮次为 20 000，α,β,γ 的值分别被设置为 0.5，0.4 和 0.1。如果一个社交网络的流行程度a_i^* 小于 0.1 或者大于 0.9，则仿真结束，并记录结束时的轮次。

7.5.2.3　仿真结果与分析

本实验中生成了一个小世界网络 SN_1 和一个均匀度网络 SN_2，并比较了两者在用户关注度方面的竞争演化过程。图 7-11 和图 7-12 为用户关注度竞争时间与两个网络背叛诱惑参数 b_1 与 b_2 之间的关系，红色区域中的结束轮次要高于蓝色区域中的结束轮次，这说明红色区域中的用户关注度演化过程要长于蓝色区域。可以看出，当小世界网络的背叛诱惑参数高于均匀度网络时，其演化时间较长；而当均匀度网络的背叛诱惑参数高于小世界网络时，其演化时间较短。这说明，均匀度网络中的 b 值对用户关注度的影响要高于小世界网络中 b 值的影响。另外，比较图 7-11 和图 7-12 可以看出，当用户对自己的声望更加关注时（p_r 值较高时），总的演化周期会普遍变长，这说明高 p_r 值对于两个社交网络的演化博弈过程具有阻滞和降温的作用。或者可以说，声望的重要程度参数同样可以影响到协同演化博弈过程：一个更大的 p_r 可以压制背叛参数 b 的影响。

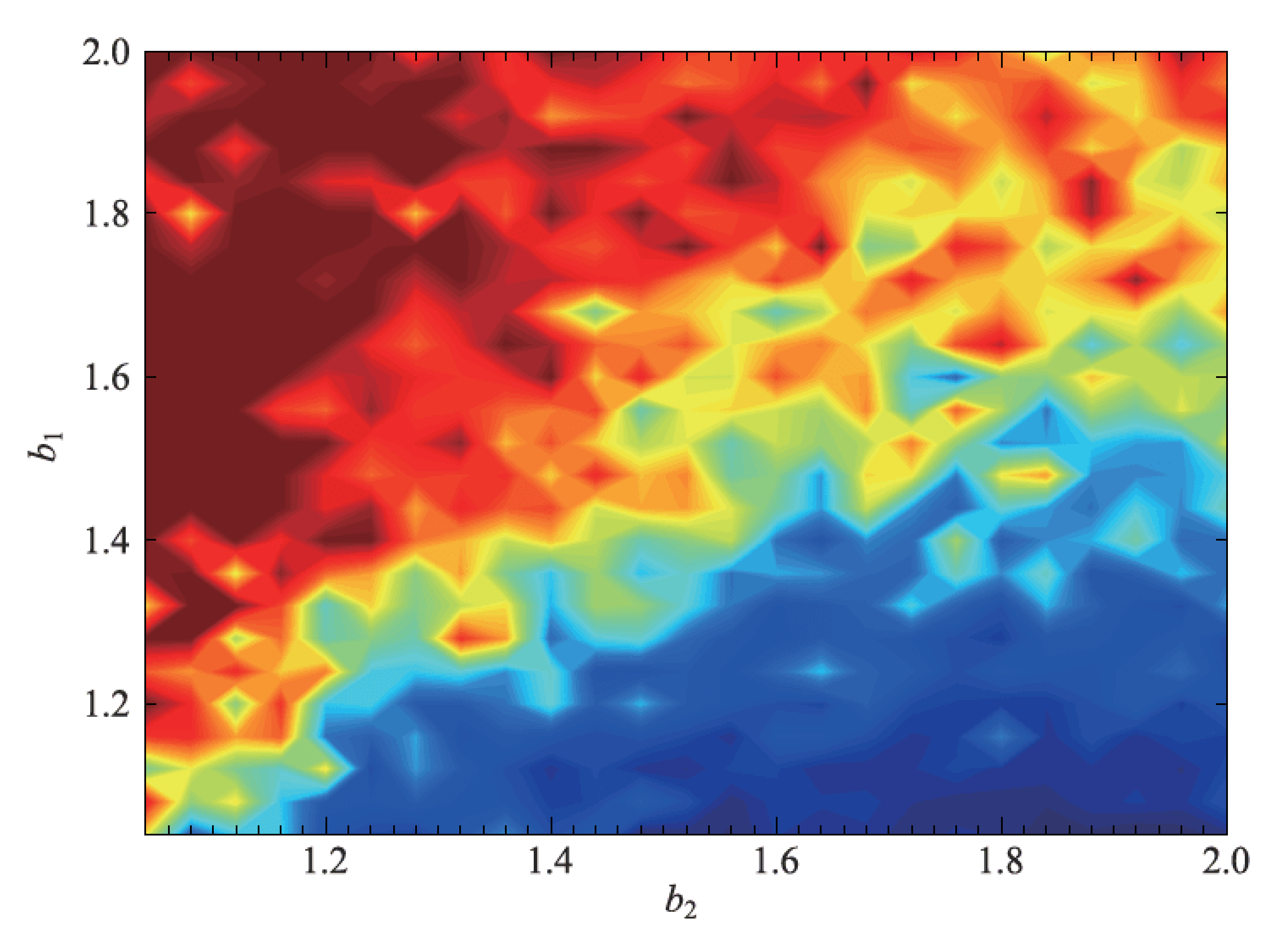


图 7-11　用户关注度随 b 的变化趋势（$p_r=0.2$）

图 7-13 和图 7-14 给出了用户关注度竞争时间与初始 p_r 值之间的关联关系，可以看出，均匀度网络的初始 p_r 值对演化时间的影响要明显大于小世界网络的初始 p_r 值，其中，p_r 值越小，演化时间越短，这说明，p_r 虽然是一个可变参数，但初始的 p_r 值对网络演化的进程具有决定性的影响。另外，通过图 7-13 与图 7-14 的对比可以确认，更高的均匀度网络中的 b_2

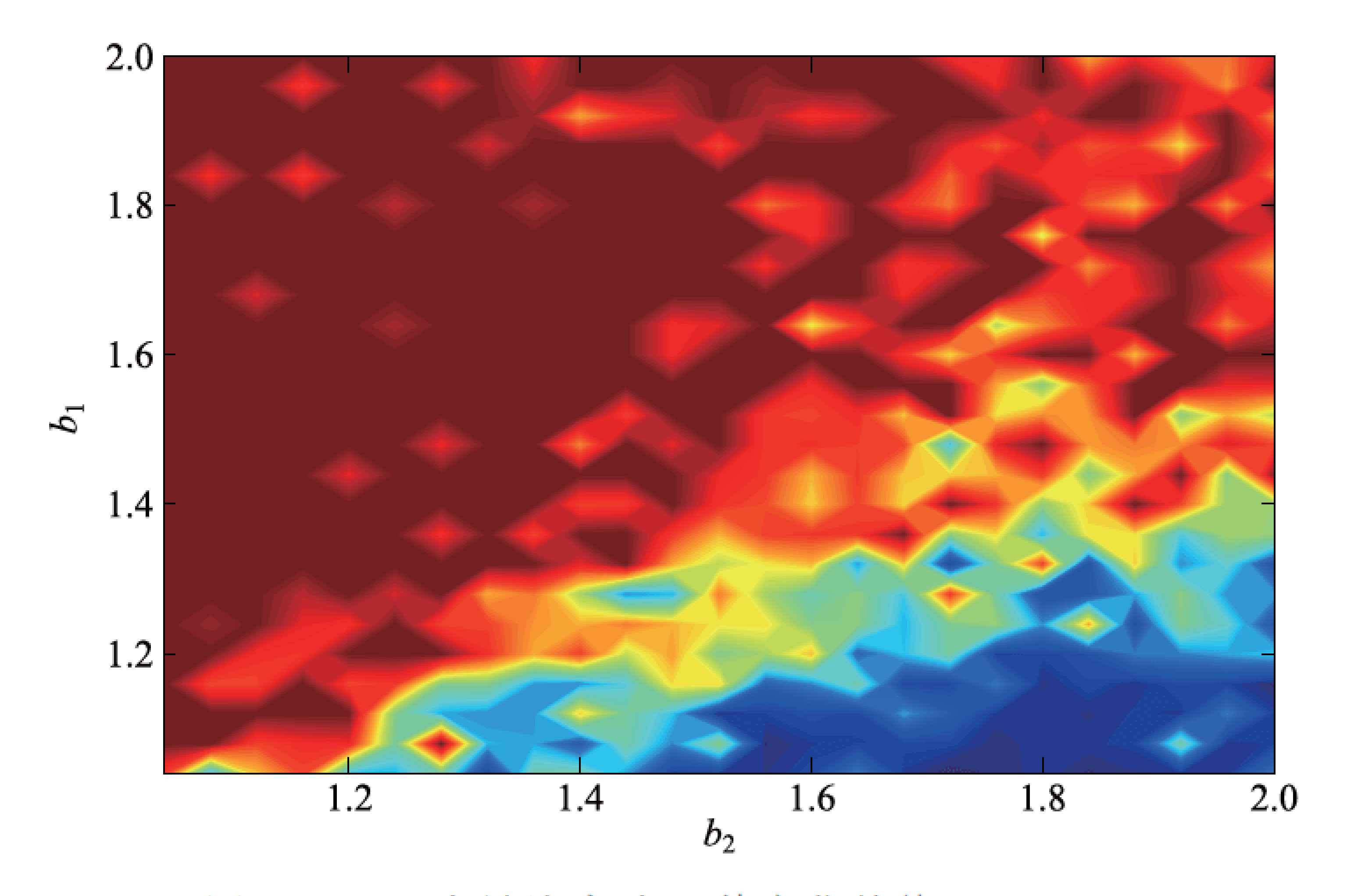


图 7-12　用户关注度随 b 的变化趋势（$p_r = 0.8$）

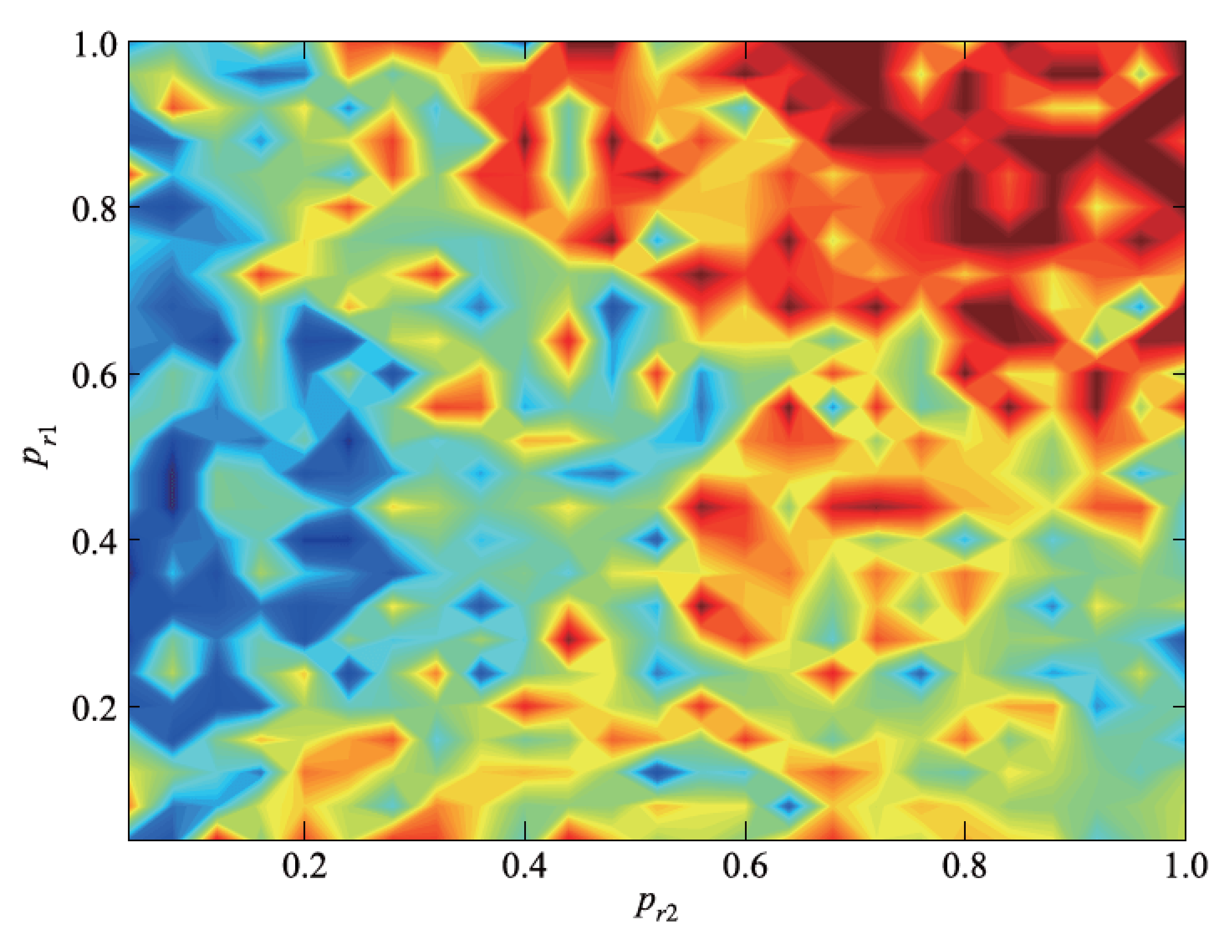


图 7-13　用户关注度随 p_r 的变化趋势（$b = 1.2$）

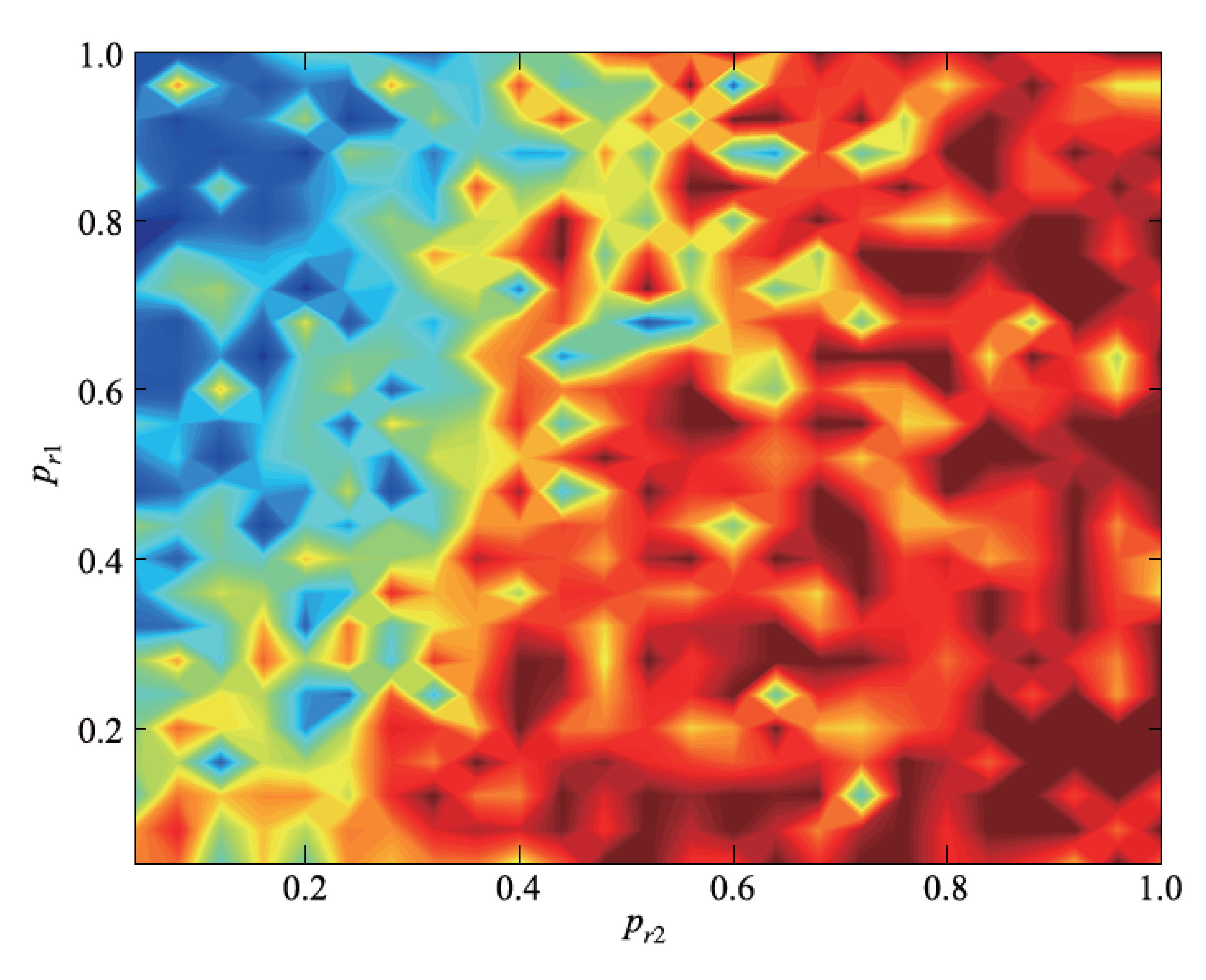


图 7-14　用户关注度随 p_r 的变化趋势（$b = 1.8$）

值会增加演化博弈的时间;这说明背叛参数在初始 p_r 值较低的时候存在动态平衡的情况，亦即：对声望的不重视导致 b 的变化对两个网络间的用户关注度竞争的作用降低，而使得关注度演化时间变长。

图 7-15 和图 7-16 给出了一组小世界与均匀度两个网络选择的初始连边密度不同时的最终演化轮次实验结果，初始连边密度 degree＝1,2,…,7,共 7 组实验。图 7-15 为最终轮次随初始 p_r 的变化趋势，可以看出，在两个网络选择的 b 值相同的情况下，结束轮次随 p_r 的增加而显著增长，且初始连边密度较低时演化时间较长。图 7-16 为最终轮次随 b 的变化趋势，可以看出，在两个网络选择的 p_r 值相同的情况下，初始连边密度以及 b 值的选择对网络的演化时间没有明显的影响。

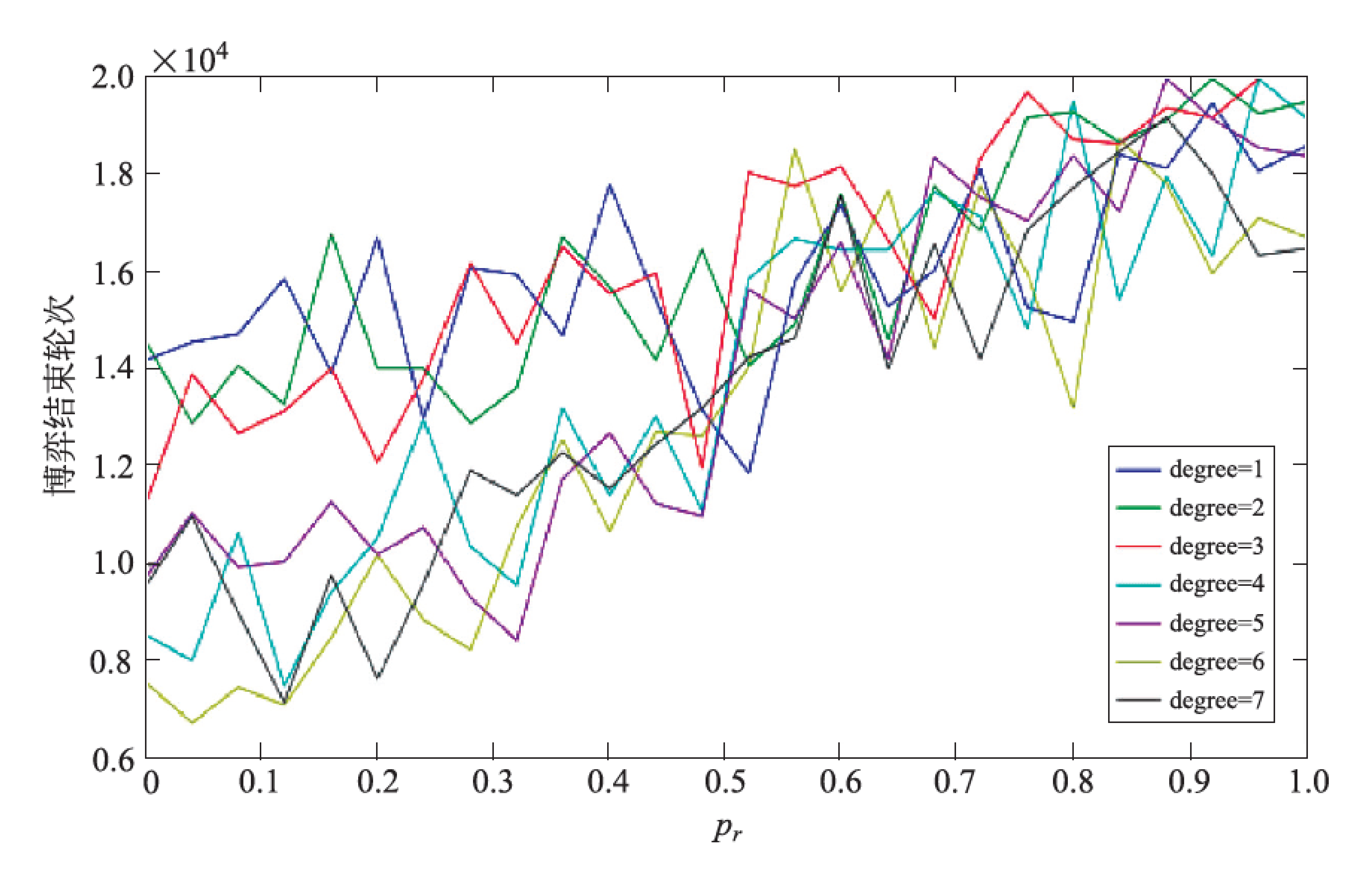


图 7-15 演化结束轮次与初始网络密度关联分析(p_r为变量)

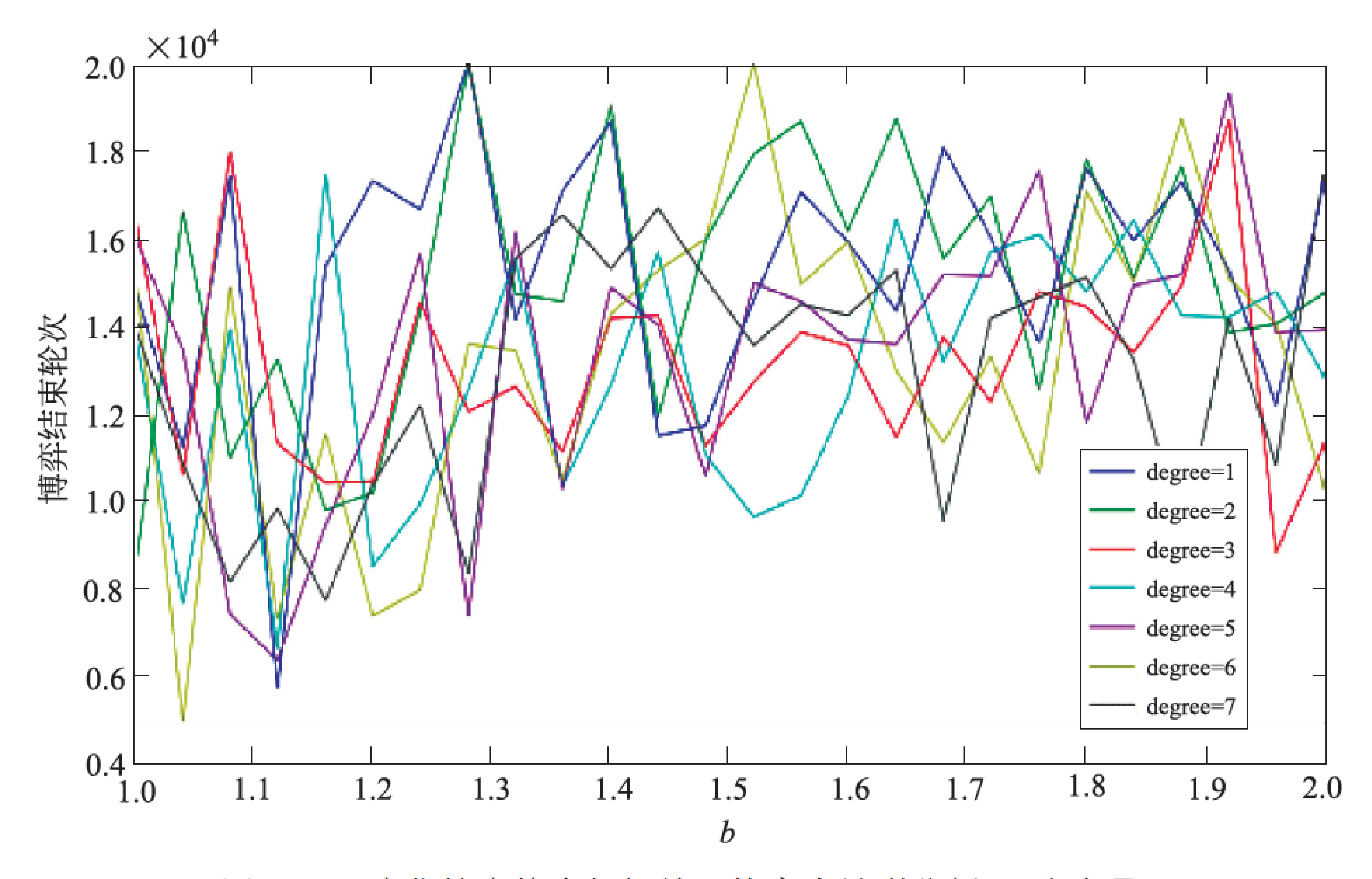


图 7-16 演化结束轮次与初始网络密度关联分析(b 为变量)

小结

在本章中还讨论社交网络中普遍存在的用户跨多个群体的现象，并在这种群体相互影响的情形下对群体上的合作演化进行分析。结果发现，在相互影响的群体中存在着网络互惠现象。在相互影响的群体中，合作个体需要更多的时间进行调整和聚集。在这一调整时期内，合作者要承受背叛者更多的入侵，但是随着个体关系更新频率的变大，合作者可以有效地抵御背叛者的入侵。当个体策略与关系发生共演化时，单个体影响与群体影响的对合作演化的影响的差别不如方格网络上的影响的差别大。因此，对于相互影响群体上个体策略和关系的共演化对合作及网络结构的影响还有待更多的研究与分析。

针对用户关注度的竞争，提出了一个竞争性社交演化博弈模型用以研究多个社交网络之间对于用户关注度的竞争过程。给出了用户关注度的概念，以及社交网络的流行程度概念，用以描述社交网络用户在社交网络服务上的时间分布。通过一组合作/背叛取值不同的均匀度网络与小世界网络的博弈模拟实验，证明了更大的声望重视程度p_r可以有效地抑制背叛收益b的影响。

参考文献

[1] Wilson C, Boe B, Sala A, *et al*. User interactions in social networks and their implications. In: Proc. of the 4th ACM European Conf. on Computer Systems. 2009, 205-218.

[2] 林闯，田立勤，王元卓. 可信网络中用户行为可信的研究. 计算机研究与发展，2008，45(12)：2033-2043.

[3] Newman MEJ. Assortative mixing in networks. Physical Review Letters, 2002, 89(20): 208701.

[4] Newman MEJ. Fast algorithm for detecting community structure in networks. Physical Review E, 2004, 69(6): 066133.

[5] Arenas A, Duch J, Fernández A, *et al*. Size reduction of complex networks preserving modularity. New Journal Physics, 2007, 9(6): 176-189.

[6] Leicht EA, Newman MEJ. Community structure in directed networks. Physical Review Letters, 2008, 100(11): 118703.

[7] Guimerà R, Sales-Pardo M, Amaral LAN. Module identification in bipartite and directed networks. Physical Review E, 2007, 76: 036102.

[8] Barber MJ. Modularity and community detection in bipartite networks. Physical Review E, 2007, 76(6): 066102.

[9] Traag VA, Bruggeman J. Community detection in networks with positive and negative links. Physical Review E, 2009, 80(3): 036115.

[10] Mucha PJ, Richardson T, Macon K, *et al*. Community structure in time-dependent, multiscale, and multiplex networks. Science, 2010, 328(5980): 876-878.

[11] Arenas A, Fernández A, Gómez S. Analysis of the structure of complex networks at different resolution levels. New Journal of Physics, 2008, 10(5): 053039.

[12] Reichardt J, Bornholdt S. Statistical mechanics of community detection. Physical Review E, 2006, 74(1): 016110.

[13] Arenas A, Fernandez A, Fortunato S, *et al*. Motif-Based communities in complex networks.

Journal of Physics A：Mathematical and Theoretical，2008，41(22)：224001.
[14] Brandes U，Delling D，Gaertler M，*et al*. On modularity clustering. IEEE Trans. on Knowledge and Data Engineering，2008，20(2)：172-188.
[15] Wang WX，Ren J，Chen GR，*et al*. Memory-Based snowdrift game on networks. Physical Review E，2006，74(5)：056113.
[16] 李峰，沈惠璋，张聪，等. 社会影响的外部性与群体动力学研究. 科技管理研究，2010，21：224-229.
[17] Chun H，Kwak H，Eom Y-H，*et al*. Comparison of online social relations in volume vs interaction：A case study of cyworld. In：Proc. of the 8th ACM SIGCOMM Conf. on Internet Measurement. 2008. 57-70.
[18] Fortunato S. Community detection in graphs. Physics Reports，2010，486(3)：75-174.
[19] Girvan M，Newman ME. Community structure in social and biological networks. Proc. of the National Academy of Sciences，2002，99(12)：7821-7826.
[20] Fortunato S，Barthelemy M. Resolution limit in community detection. Proc. of the National Academy of Sciences，2007，104(1)：36-41.
[21] Newman ME，Leicht EA. Mixture models and exploratory analysis in networks. Proc. of the National Academy of Sciences，2007，104(23)：9564-9569.
[22] Copic J，Jackson MO，Kirman A. Identifying community structures from network data via maximum likelihood methods. The BE Journal of Theoretical Economics，2009，9(1).
[23] Doreian P，Batagelj V，Ferligoj A. Generalized Block Modeling. Cambridge University Press，2005.
[24] Zhou D，Manavoglu E，Li J，*et al*. Probabilistic models for discovering e-communities. In：Proc. of the 15th Int'l Conf. on World Wide Web. 2006. 173-182.
[25] Pathak N，Delong C，Banerjee A，*et al*. Social topic models for community extraction. In：Proc. of the 2nd SNA-KDD Workshop，Vol. 8. 2008.
[26] Sachan M，Contractor D，Faruquie TA，*et al*. Using content and interactions for discovering communities in social networks. In：Proc. of the 21st Int'l Conf. on World Wide Web. 2012. 331-340.
[27] Qi GJ，Aggarwal CC，Huang T. Community detection with edge content in social media networks. In：Proc. of the 2012 IEEE 28th Int'l Conf. on Data Engineering (ICDE). IEEE，2012. 534-545.
[28] Chen W，Liu Z，Sun X，*et al*. A game-theoretic framework to identify overlapping communities in social networks. Data Mining and Knowledge Discovery，2010，21(2)：224-240.
[29] Palla G，Derényi I，Farkas I，*et al*. Uncovering the overlapping community structure of complex networks in nature and society. Nature，2005，435(7043)：814-818.
[30] Alvari H，Hashemi S，Hamzeh A. Discovering overlapping communities in social networks：A novel game-theoretic approach. AI Communications，2013，26：161-177.
[31] Narayanam R，Narahari Y. A game theory inspired，decentralized，local information based algorithm for community detection in social graphs. In：Proc. of the 21st Int'l Conf. on Pattern Recognition (ICPR 2012). 2012. 1072-1075.
[32] Basu S，Maulik U. Community detection based on strong nash stable graph partition. Social Network Analysis and Mining，2015，5(1)：1-15.
[33] Jiang F，Xu J. Dynamic community detection based on game theory in social networks. In：Proc. of the 2015 IEEE Int'l Conf. on Big Data (Big Data). IEEE，2015. 2368-2373.
[34] Zhan W，Guan J，Chen H，*et al*. Identifying overlapping communities in networks using evolutionary method. Physica A：Statistical Mechanics and its Applications，2016，442：182-192.
[35] Alós-ferrer C，Ania AB. Local equilibria in economic games. Economics Letters，2001，70(2)：165-173.

[36] Shen H, Cheng X, Cai K, *et al*. Detect overlapping and hierarchical community structure in networks. Physica A: Statistical Mechanics and its Applications, 2009, 388(8): 1706-1712.

[37] Monderer D, Shapley LS. Potential games. Games and Economic Behavior, 1996, 14(1): 124-143.

[38] Hoffman MD, Blei DM, Bach FR. Online learning for latent dirichlet allocation. In: NIPS, Vol. 2. 2010. 5.

[39] Facebook Statistics. 2016. https://facebook.com/press/info.php? statistics.

[40] Yu J, Wang Y, Jin X, *et al*. Evolutionary analysis on online social networks using a social evolutionary game. In: Proc. of the 23rd Int'l Conf. on World Wide Web, WWW'14. New York, 2014. 415-416.

[41] Yu J, Wang Y, Jin X, *et al*. Social evolutionary games. In: Proc. of the 5th Int'l Conf. on Game Theory for Networks, GAMENETS. Beijing, 2014. 1-5.

[42] Yazicioglu AY, Ma X, Altunbasak Y. Evolution of mixed strategies for social dilemmas on structured networks. In: Proc. of the 2011 Int'l Conf. on Networking, Sensing and Control. Delft, 2011. 175-180.

[43] Wu L, Chen X. Modeling of evolutionary game between SNS and user: From the perspective of privacy concerns. In: Proc. of the 21st Int'l Conf. on Management Science & Engineering. Helsinki, 2014. 115-119.

第三部分

跨网络的社交网络演化计算

第二部分给出了单一网络的社交网络演化计算模型、方法和实施用例。本部分在第二部分的基础上进一步讨论跨网络的社交网络演化计算问题,包括跨网络的可预测性分析、跨网络强关联账户(异构网络中的多个账户确定属于同一个自然人或团体)之间的用户行为关联预测,以及弱关联账户(异构网络中的多个账户属于具有相似特征的人或团体)之间的用户行为偏好关联分析等。

第 8 章给出跨网络的可预测性分析。首先给出用户特征模型、用户特征建模学习方法以及用户特征表述方式的经典研究成果,随后基于在线社交网络的用户数据,提出一种基于知识库理解的用户时空特征建模方法(TUMK),用以学习用户特征。该方法试图将用户在社交网络和消费网络中的显式反馈和隐式反馈特征转变成机器可识别理解的向量,针对用户反馈信息类型构建差异化的特征建模方法。特别是针对用户在社交网络中发布的文本类消息,利用知识库概念提出基于统计模型的实体识别与概念标注一体化过程。

进一步地,第 9 章给出跨网络的强关联账户之间的用户行为关联预测分析,首先简介个性化的预测技术、预测方法以及推荐系统的评价指标体系,随后基于第 8 章中学习的用户特征,提出一种融合异质社交网络的消费媒体用户偏好预测方法(FHSP),将目标用户在消费媒体中的特征、社交网络特征与消费媒体中的产品特征综合建模,基于协同矩阵分解思想,利用梯度下降学习最优参数,有效地融合多平台资源,对用户偏好做出预测。

更进一步地,第 10 章给出跨网络弱关联账户的用户行为偏好关联分析,基于用户在社交网络和消费媒体中的显式特征和部分隐式特征,扩展博弈论稳定匹配算法,提出基于博弈思想的异质网络账户匹配方法(HAM),并在 Facebook 和 eBay 数据集上证明,在显式特征数据稀疏的情况下,本书的算法可以避免长尾效应,有效匹配异质网络账户。

第 8 章 跨网络用户偏好可预测性分析

用户特征建模是指从能够体现用户个性化的信息（显式反馈和隐式反馈）中，归纳出可计算的用户特征模型的过程。用户特征建模不断收集和处理用户数据，以得到用户的一些特性，比如兴趣、知识、目标以及其他一些事实。从形式化描述的角度上看，用户特征建模是一种面向算法的、具有特定数据结构的、形式化的对用户兴趣偏好的一种数学描述。详细、准确的用户特征建模是提供优质的个性化服务的基础和前提，一个结构合理、描述准确的用户兴趣模型，可以为个性化服务提供详实、易操作的用户个性化信息。用户特征作为个性化服务的基础和核心，直接关系到个性化服务的质量。本章将学习已有的用户特征建模方法，结合用户在社交网络和消费网络的信息，针对异质网络推荐应用，提出一种新的面向大规模用户数据集、利用知识库、快速准确实时的用户特征建模模型 TUMK（temporal user model based on knowledge base）。

8.1 概述

当用户加入一个线上服务，特别是社交网络服务时，往往需要提交一些基本的个人信息，比如昵称、年龄、性别等，部分网站甚至要求用户填写一段自我简介以方便服务商和其他用户的社交关系识别。这些信息通常称为显式反馈（explicit feedback），是用户的一个特征表述。用户的显式反馈是用户特征非常重要的一部分，对应具体应用具有关键性作用，比如 Yu 等人[1]的研究表明，使用用户资料中的人文特征信息对诸如网页广告投放等应用具有非常重要的指导意义，进而设计出一种基于大规模人文特征标示的用户特征建模方法。

同时，当用户使用一个在线社交网络或在线消费网络时，用户会在这些服务之上留下很多足迹，比如用户在社交网络中发布信息、与好友

互动,在购物媒体中购买感兴趣产品并留下评价。这些信息通常称为隐式反馈(implicit feedback),是用户特征的一种行为表现。随着大数据时代的来临,研究人员对短文本、非正式语言的理解有了更多的处理手段,利用用户隐式反馈对用户个性化建模具备了可行性,如Baltrunas的研究工作[2]表明,在融合多方法的推荐系统中,用户的隐式反馈对推荐准确性越来越重要。

更进一步地,本章基于前人研究成果,针对用户在社交网络和消费网络中的显式反馈和隐式反馈进行特征建模,特别是隐式反馈,提出基于知识库理解的用户特征建模。从维度上,可以从空间维度和时间维度进行刻画:时间维度上,认为用户兴趣偏好是随时间延伸而变化的,针对某一个时间点的用户特征设法使用距目标时间最近的用户特征;而在空间维度上,主要针对用户数据稀疏问题,试图融合目标用户好友或相似用户的特征给出目标用户特征。

8.2 相关研究工作

个性化服务是Internet信息增长的必然结果。与不区分用户的普通服务模式不同,个性化服务具有更高的服务质量,可以依据用户特性,以恰当的方式为用户提供恰当的服务。优秀的个性化服务可以为用户提供良好的体验,并提高服务商的收益。然而,详细、准确的用户兴趣模型是提供优质的个性化服务的基础和前提,一个结构合理、描述准确的用户兴趣模型,可以为个性化服务提供详实、易操作的用户个性化信息。用户特征作为个性化服务的基础和核心,直接关系到个性化服务的质量。近几年来,越来越多的学者认识到准确的用户特征在应用中的重要性,并做了诸多相关研究工作。Fragoudis对典型的个性化服务系统LIRA等采用的用户兴趣建模方法进行了综述和分析,指出用户兴趣建模在个性化服务系统中的重要地位[3]。如图8-1所示,本节分别从用户特征模型、建模过程以及表达方式这3个方面叙述用户特征建模的相关内容。

8.2.1 用户特征模型

用户特征模型主要可以分为4类:静态模型(static model)[4]、动态模型(dynamic model)[5,6]、类型模型(stereotype model)[7,8]和自适应模型(adaptive model)[9]。

静态模型是指用户的信息一旦被收集并构建成特征模型,将一直被应用而不会改变。该模型作为最常用的一类模型,其优点是简单、易用,且计算复杂性低。通常,静态模型的用户特征仅使用几类用户显式信息,比如用户性别,或者用户一开始登录时提交的一些偏好,而不能学习处理历史行为等隐式反馈特征。因此,静态模型较为简单,并且为了保证模型的简洁性,一旦建立就不易改变,因此在跟踪用户偏好变化方面存在缺陷。

动态模型与静态模型恰恰相反,允许用户特征加入时间维度,用户特征会随用户的新的显式反馈和隐式反馈更新而变化,因此会随时间推进依用户反馈变化而变化。Liliana[5]构建了一个实时收集用户信息的系统,根据用户隐式反馈不断调整用户特征,其结果证明,用户特征会发生显著变化。动态模型更适合用户特征学习。Ahmed提出了TVUM模型[6],同时考虑用户长期和短期的兴趣特征,实验结果表明,这种模型在随机性较大的广告点击预

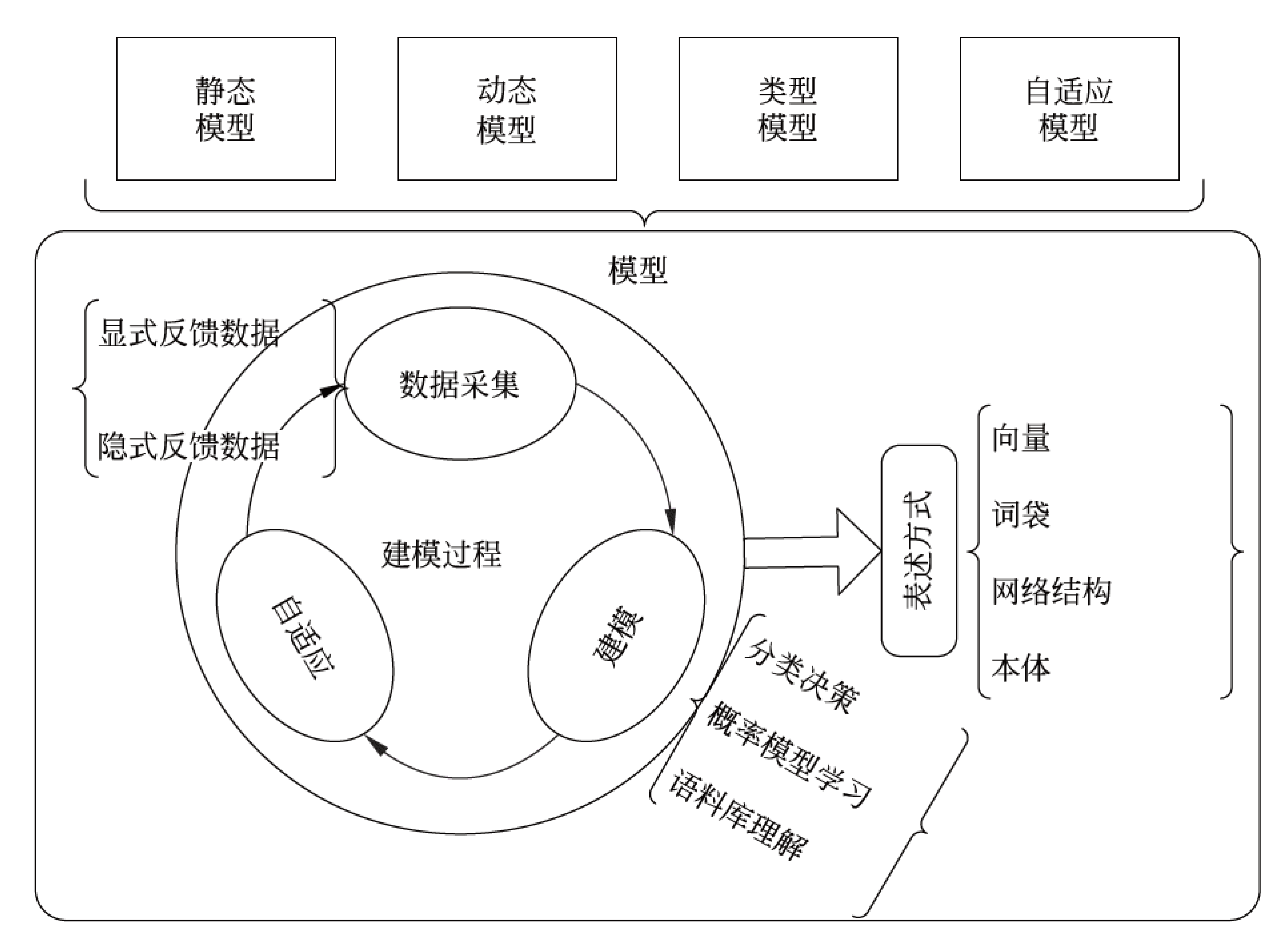


图 8-1　用户特征建模组成

测中效果明显。动态模型考虑到用户偏好随时间变化的特征，但是用户的反馈呈现长尾分布，大多数用户反馈数据较稀疏，准确计算用户偏好或特征存在一定困难。

为了降低数据稀疏无法准确学习用户特征这一情况，类型模型首先将用户划分到一些预先指定的类型组中，其假设是相同特征的用户具有相似的特征模型。如对同组用户构建特征，并赋给该类别所有的用户，使这些用户共享该特征。类型模型最早由用户数据分析大师 Rich[4] 提出，他认为原型模型在用户只有少量信息时，是最有效的用户模型。Albanil[7] 认为，原型模型特别适用于用户划分的场景，并提出一种原型模型，应用在医疗咨询中，提高了医学检测的准确度。Kobsa[8] 回顾了基于原型的用户特征建模，之后，进一步分析了基于原型的方法的自适应性：基于原型的用户特征建模，使用用户常用的关键词来表示，并具有自适应性。然而，原型是计算的一类用户的特征，用户个性化在一定程度上会被损耗，尤其是在划分不合适的情况下，用户个性化特征将会消失，可能会带来 one-size-fits-all 问题[9]。

自适应用户特征建模模型如图 8-2 所示。

自适应模型依据具体应用场景，不断依据应用计算结果调整用户特征学习过程，保证每个用户数据以最合适的方式学习。该模型不仅从数据出发，同时考虑到应用的反馈，有针对性地对用户建模。由于高性能的计算和存储资源越来越容易获得，自适应模型正逐渐代替动态模型和静态模型。Lee 等人[10] 根据推荐结果，选择性地使用用户数据用以构建用户特征，从而有效促进了推荐准确度。但是，该模型将用户特征建模与具体应用绑定，不具备通用性且往往计算复杂性较高，分析复杂。

8.2.2　用户特征建模学习方法

从图 8-1 可以看出，用户建模通常可以分为 3 个独立的阶段：数据采集（data collection）、用户模型计算（modeling）和自适应（adaptive）[4]。其中，数据采集阶段收集尽可

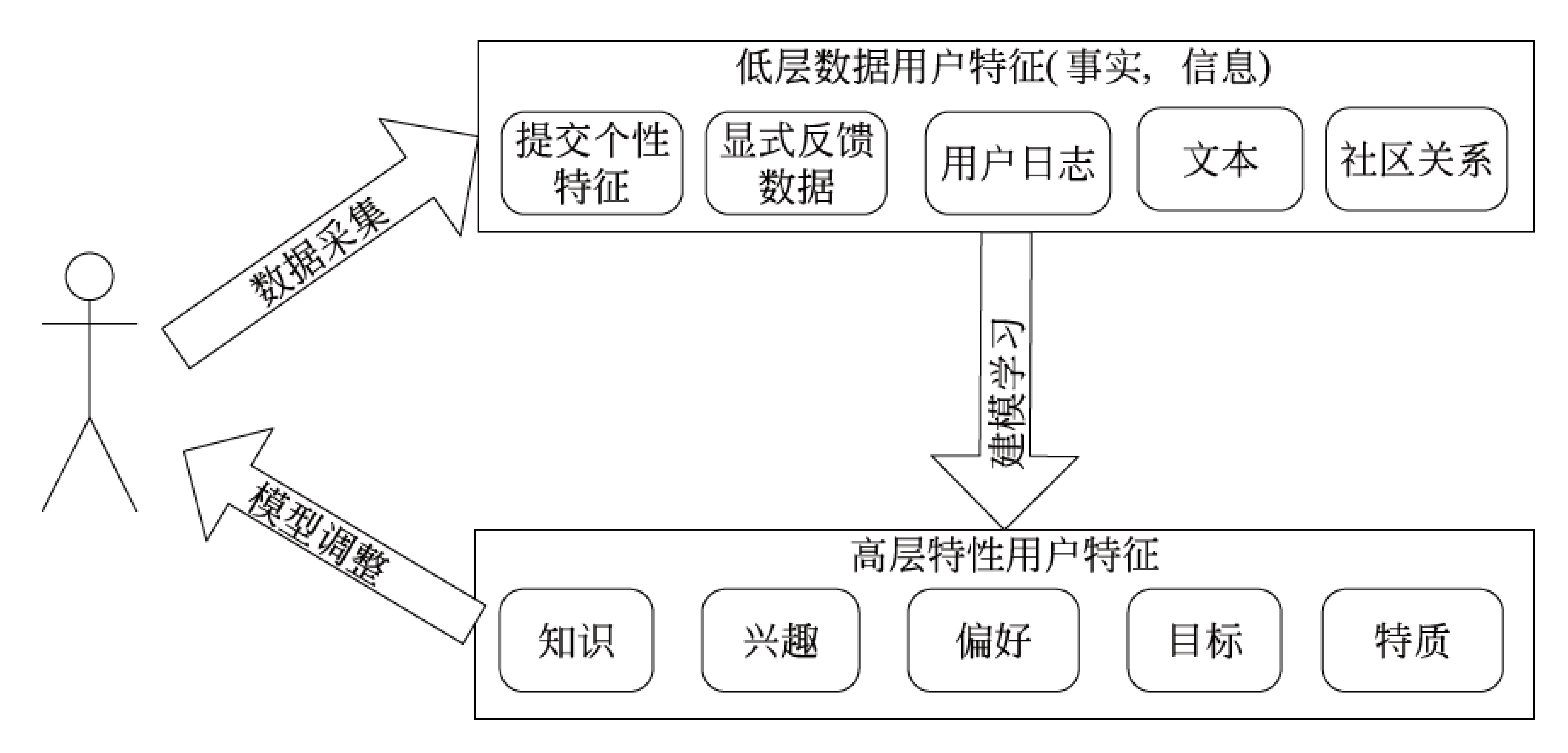


图 8-2　自适应用户特征建模模型

能多的用户的各式数据和信息，所有的用户反馈都作为用户的一些事实数据加以存储。数据采集阶段将作为数据链路层的第 1 阶段，供后续数据选择计算。用户建模阶段作为数据链路层的第 2 阶段，试图将直接存储的用户反馈通过学习计算转化为机器可以理解的向量等表示方法，例如用户兴趣、偏好或者用户其他特性。在该阶段，用户模型只是通过反馈学习而得到的。自适应阶段是针对具体应用优化用户特征模型，通过具体应用对用户特征反馈，可以更准确地掌握用户特征。

其中，在用户模型学习阶段，可以基于不同形式的用户反馈数据使用相应的优化学习方法。主要可以从两方面数据来描述：显式反馈(explicit feedback)和隐式反馈(implicit feedback)[11,12]。显式反馈是指直接询问用户的兴趣和偏好、允许用户自己定义和修改用户模型来构建用户兴趣模型，特点是信息简洁、描述精确、计算复杂度低；而隐式反馈是指需要跟踪用户的浏览行为和交互行为等信息，隐式信息数据量大，内容丰富，但是特征提取相对复杂。针对用户显式反馈和隐式反馈，通常有 3 类学习方法：分类决策(classification)[13,14]、模型学习(model learning)[15]和语料库学习(corpus learning)。

分类决策方法是指事先依据专家经验，对实体进行分类，构建分类树；根据用户显式信息和隐式反馈，将用户逐渐分类到具体一类特征。Lakiotaki 等人[13]通过学习现实世界的分类科学，构建多层次分类模型，对用户反馈进行分类并应用于推荐系统中。Iglesias[14]认为，用户信息可以转化成一系列的行为或指令，因此可以构建一类类似 trie 的分类树，对用户连续行为进行学习并经实验证明在特定场景下效果显著。分类决策方法计算简单、复杂性低，但决策树构建难度大，需要大量数据标注分类，而且不具备动态性，无法适应改变。

模型学习利用一些概率统计模型，比如隐语义模型，学习文本主题，并将主题赋予为用户特征。Rossignol 等人[15]为了学习用户文本消息所表达的信息，提出一种基于朴素贝叶斯的用户对话演化模型，利用最大后验概率学习参数，经实验，其结果表明可以准确学习用户对话内容，特别是在对话数据缺失的情况下。模型学习不需要提前构建语料，而是根据学习语境自适应学习，有着较高的准确度。但当面临大量数据时，由于计算复杂性高，且对于一些诸如短文本等数据缺少分析手段，如果不依赖外部语料，很难学习其中的实体与概念。

语料库学习是近几年来比较流行的方法，通过借助构建好的语料库以及用户反馈中的实体，将其概念作为用户特征。Abel 等人[16]利用开放语料库 OpenCalais，针对 Twitter 用户上的信息抽取实体和主题，以更好地理解用户表达的意思或参与的活动。Gao 等人[17]提

出 GeniUS 模型,利用外部构建好的知识库理解用户特征,可以成功推断出特殊领域的概念知识。基于外部语料库的学习简单、高效,特别是语料库越丰富,解析效果越优秀。但是,一旦语料数量和内容丰富但分布严重不均时,如果不辅以概率模型理解,将不能深入理解语言内部的含义,而只能给出表象特征。

8.2.3 用户特征表述方式

对于用户的特征表述方式,通常有 4 类形式:基于向量标签(vector based)[18,19]的表述方式、基于词袋(bag of words)[20]的表述方式、基于网络结构(network based)的表述方式以及基于本体(ontology based)[21,22]的表述方式。

向量标签通常用来表示实体的特征、分类或内容,通常以关键词的形式给出。基于标签的用户特征建模,就是对用户的信息以关键词或者特征向量的形式表示。对于显式信息,比如用户昵称、年龄、性别、偏好等,可以直接用标签的形式给出;对于隐式信息,比如用户历史行为,可以通过分类、归类等打上标签。这种方式简单、实用,并具有普遍性,在用户特征建模研究早期是一种常见的建模方式,如 Carmagnola 等人[18]将用户的个人资料使用标签表示,并确认其对个性化推荐的准确性有促进作用。除了直接将用户特征直接表示以外,也可以对用户特征做一些泛化,比如对于用户的年龄,Liu[19]的工作中就将 age 属性转变成 ageRange 属性。基于向量标签表示的用户特征简单、明确,但是使用范围较小,仅适用于一些静态特征学习表示,不适用于多数隐式反馈信息学习表示,也不能表达实体与实体之间的关系。

词袋模型是自然语言处理和信息检索中的一种简单假设。在用户隐式反馈数据中,特别是一些文本数据,使用词袋模型可以有效地表述用户特征。Schwarzkopf 等人[20]将学习到的用户标签以词袋形式存储,并证明词袋模型是一种适应性比较高的表达方式。词袋模型的表达方式具有普适性,但数据作为无序的词汇集合丢失了语法甚至单词的顺序,这在语法和词序通常是非常重要的特征的自然语言理解问题中是一个较大的损失。

基于网络表述的方式将用户模型表述成一种语义网络,其中,节点表示概念,边表示概念之间的关系。相对于词袋模型,基于网络结构的表述方式不仅利用关键词表示用户兴趣,从而避免多义性,还可以有效地利用其边,即关联性,将用户特征扩展成一系列概念。

本体论(ontology)是从哲学领域借鉴过来的术语,原意是指一种存在的系统化解释,借鉴至知识工程领域之后,研究者对本体论进行了新的诠释,认为本体论是概念化对象的明确表示和描述。由于本体论对特定领域对象的表示与描述具有规范性、可重用性、可靠性等特点,研究者将本体论应用于检索领域,对文档、用户模型进行描述,以提高系统的联想能力和精确性。本体论提出之后,大量学者试图通过本体的方式理解用户。Razmerita 等人[21]展示了一张通用的基于本体的用户特征建模机制,利用强大的知识数据系统可以有效地表达用户特征。Skillen 等人[22]试图以本体理解用户文本特征,分析了用户行为以及文本理解过程,并强调使用本体模型对用户动态构建。该表述方式形象、清晰,应用性较高,但是构建相对复杂。

8.3 基于知识库理解的时空用户特征建模

作者试图将用户无规律、无关系的信息转变成机器可理解的、可计算的特征向量。本节将充分考虑如何利用用户隐式反馈和显式反馈构建用户特征,特别是用户隐式反馈,我们尝

试使用外部语料库理解用户特征。在此基础上，考虑用户兴趣特征随时间偏移而偏移的情况，以及用户数据稀疏问题，从时间和空间维度两方面，提出基于知识库理解的用户时空特征模型(temporal user model based on knowledge base，简称 TUMK)。该模型的主要流程图如图 8-3 所示。

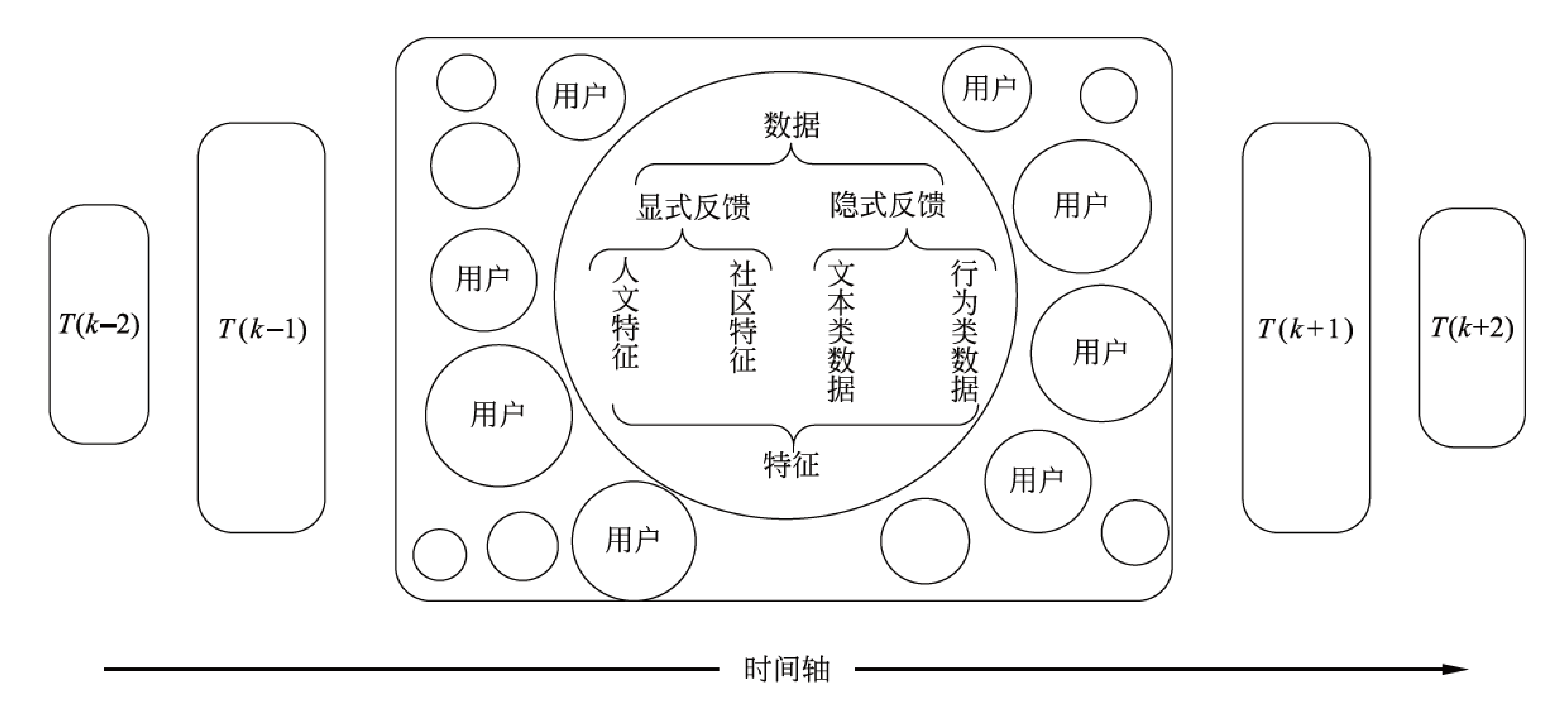


图 8-3 用户特征建模示例图

如图 8-1 所示，从用户两大类反馈，4 种不同类型的数据，给出计算方法，学习用户特征。在空间上，融合相似用户信息；在时间上，融合最近一段时间的用户特征。以下针对不同类型的数据给出详细的计算方法，并最后融合用户特征，给出目标用户在目标时间的特征。

8.3.1 基于显式反馈的用户特征建模

大量研究工作证明，用户的显式反馈在个性化服务中起着非常关键的作用。Shepstone 等人[23]发现，使用用户年龄和性别的个性化广告推荐，点击率要远远高于随机推送的广告。Kau 等人[24]发现，在线购物者通常比较年轻，擅长比较产品特性的常为年轻女性，线下购物者通常年龄超过 40 岁。而显式反馈通常包含两种类型的信息：人文特征信息(demographical)和社区特征信息(community)。人文特征信息主要针对用户、性别、地理位置等特征，而社区特征针对用户的好友关系、交互关系等图结构特征信息。

假设用户u_i的人文特征表示为d_j，首先统计所有可能的特征区间k，建立全维空间D，针对每一个用户u_i，赋予一个维度空间向量D_i，对于用户提交的特征，可以根据其数据特征规整数据。比如用户年龄(age)和地理位置(location)，将其泛化为年龄区间(age_range)和地理区间(local_range)。对于没有提交的，给予默认值为空或者平均值。基于人文特征，用户特征可以表示为如下形式：

$$F_D(u)=\{d_1,d_2,\cdots,d_k\} \tag{8-1}$$

随着网站服务性变强，多数网站提供用户添加好友关系，用户可以与好友交流互动。其中，好友对用户的影响是很大的，一个用户的好友是谁决定了用户是怎样的人。大量证据显

示，相比于陌生人或者匿名的推荐，用户更倾向于接受好友的推荐。因此，构建图结构描述用户社区数据特征。在这里图节点是用户，边代表用户好友关系。

$$F_G(u)=\{u_1,u_2,\cdots,u_m\} \tag{8-2}$$

8.3.2 基于隐式反馈的用户特征建模

用户显式反馈信息准确、易用，但数据稀疏，描述用户特征还需要更多的信息。用户隐式反馈数据量大，信息量丰富，可以深度挖掘用户的潜在特征，但用户隐式反馈具有无序性和无结构性，通常不能被计算机理解，需要将其转化成机器可理解的向量表示。用户隐式反馈也大致可以分为两类：文本类数据（text content）和行为类数据（behavior content）。文本类数据主要是指用户的一些短文本记录，比如用户在社交网络中的说话或消费网络中的评价，均可作为文本类数据；而用户点赞、收藏、转发等行为，均可视为行为类数据。

用户文本类数据，特别是社交网络中的文本类数据，其语言的无结构性和非正式性通常使机器难以理解，最常用的方法是将文本转化为固定长度的词袋模式。但是词袋模型面临两大问题，一方面，词袋模型无法利用文本中的词序意义，比如短文本"the big bang"，词袋模型中仅仅是 3 个孤立单词，其表达意义与"bang big the"完全相同；另一方面，词袋模型无法理解词义，如认为"football"和"soccer"是完全不同的概念。为解决上述问题，我们利用外部语料知识库，提出基于统计的实体与概念标注一体化模型。知识库是一个集实体、概念、属性为一体的知识数据库，其中每一个实体拥有多个属性，以不同的权重对应多种概念。每一个概念也拥有多个属性和实体，以无向图的网络结构表示。

假设用户隐式信息中文本信息表示为 $T=\langle T_1,T_2,\cdots,T_n\rangle$，其中每个文本 T_i 由实体 e 组成，$E_i=e_1e_2\cdots e_m$，实体 e_j 的概念可能为 c，即文本 T_i 是由概念序列组成，可表示为 $C_i=c_1c_2\cdots c_m$。那么，实体识别和概念标注任务就是要在 T_i 所对应的实体切分和标注形式中，寻找 E 和 C 的联合概率 $P(E,C)$ 为最优的实体切分和概念标注组合。因此，我们分别提出基于概念的统计模型和基于实体的统计模型，并在此基础上融合概念和实体模型，提出概念与实体标注一体化模型。

基于概念的统计模型如图 8-4 所示，利用隐马尔可夫链描述一个完整句子概念的变化，每种概念对应一种特征，特征之间的距离代表概念之间的搭配关系。这样，在生成一个句子时，系统不断转移概念直到句子完成。因此，$P(E,C)$ 可以由隐马尔可夫近似表示为

$$P(E,C)=P(E|C)\times P(C)=\prod_{i=1}^{n}P(e_i|c_i)\times P(c_i|c_{i-1}\,c_{i-2}\cdots c_1) \tag{8-3}$$

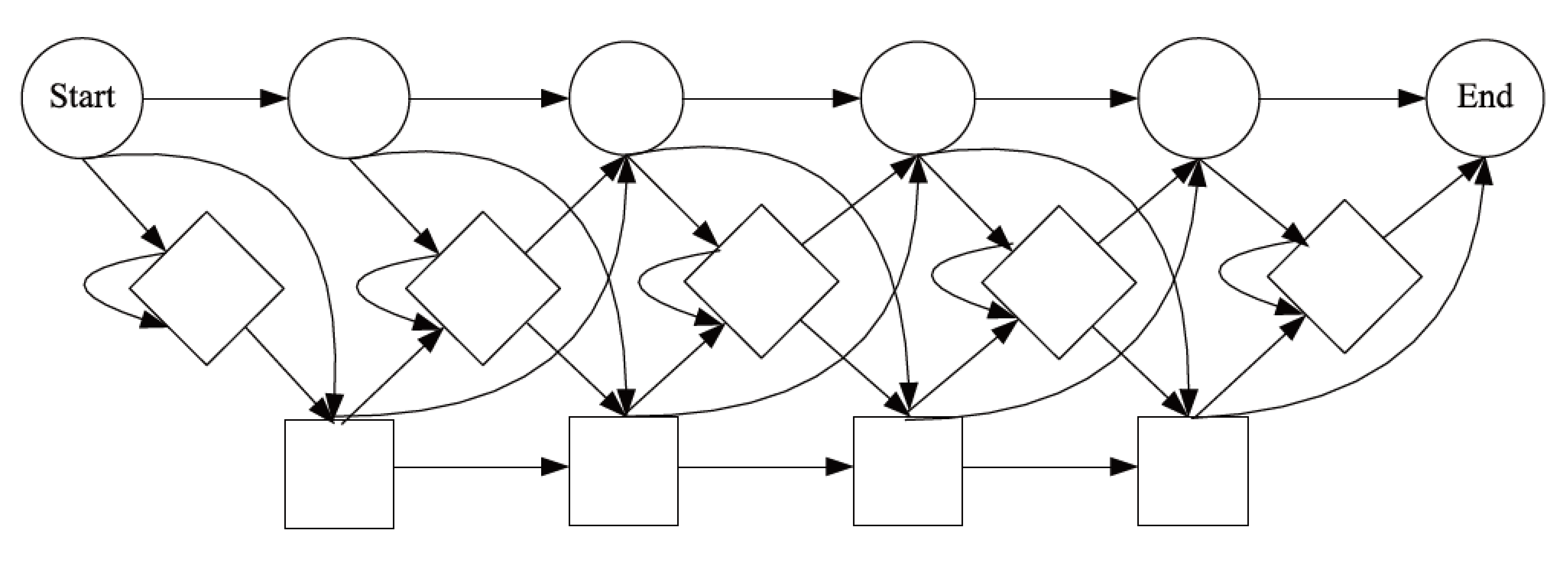


图 8-4　用户特征学习隐马尔可夫流程图

$$P(e_i|c_i)=\frac{W(e_i,c_i)}{W(c_i)} \tag{8-4}$$

$$P(c_i|c_{i-1}\ c_{i-2}\ c_{i-3})=\frac{N(c_i\ c_{i-1}c_{i-2}c_{i-3})}{N(c_{i-1}c_{i-2}c_{i-3})} \tag{8-5}$$

其中，$P(e_i|c_i)$表示在概念c_i下实体e_i出现的概率，$W(e_i,c_i)$为实体e_i对应概念c_i的权重，$W(c_i)$为概念c_i所有权重之和。$P(c_i|c_{i-1})$表示从概念c_{i-1}到概念c_i的转移概率，可以利用知识库中概念之间的转移概率来计算。$N(c_ic_{i-1}c_{i-2}c_{i-3})$表示概念$c_i$、$c_{i-1}$、$c_{i-2}$、$c_{i-3}$所共同对应实体的数量。

同样地，如果把实体作为中间状态，概念作为输出，该模型称为基于实体的统计模型，则$P(E,C)$的另一种表示方式是

$$P(E,C)=P(C|E)\times P(E)=\prod_{i=1}^{n}P(c_i|e_i)\times P(e_i|e_{i-1}e_{i-2}\cdots e_1) \tag{8-6}$$

$$P(c_i|e_i)=\frac{W(e_i,c_i)}{W(e_i)} \tag{8-7}$$

$$P(e_i|e_{i-1}e_{i-2}e_{i-3})=\frac{N(e_ie_{i-1}e_{i-2}e_{i-3})}{N(e_{i-1}e_{i-2}e_{i-3})} \tag{8-8}$$

其中，$P(c_i|e_i)$表示实体c_i的所有概念中，e_i所占比例。$W(e_i,c_i)$为实体e_i对应概念c_i的权重，$W(e_i)$为实体对应所有概念的权重之和。$P(e_i|e_{i-1}e_{i-2}\cdots e_1)$表示实体之间的转移概率，$N(e_ie_{i-1}e_{i-2}e_{i-3})$表示实体$e_i$、$e_{i-1}$、$e_{i-2}$、$e_{i-3}$所共同对应概念的数量。

综合上述模型，可以给出一种实体与概念标注一体化模型，并可通过调整参数α和β，优化计算结果以提高准确度。

$$\hat{P}(E,C)=\alpha\prod_{i=1}^{n}P(e_i|c_i)\times P(c_i|c_{i-1}c_{i-2}\cdots c_1)+\beta\prod_{i=1}^{n}P(c_i\mid e_i)\times P(e_i|e_{i-1}e_{i-2}\cdots e_1) \tag{8-9}$$

基于概念与实体标注一体化模型，可将学习的概念作为用户特征。不同于普通词袋表示，学习概念带有权重，表示用户对该概念的偏好程度，其特征可以表示为

$$F_T(u)=\{c_1:w_1,c_2:w_2,\cdots,c_k:w_k\} \tag{8-10}$$

对于用户行为类数据，如果用户关注某项产品，则可以认为用户喜好该类成品。那么假设产品表示为e，在消费媒体中，每一项产品都从属于一类，比如eBay将所有商品划分为36类、YouTube将视频划分为17类。通过统计用户u所有点击过的产品$[e_1,e_2,\cdots,e_n]$，统计每类产品对应类别$C(e_1)$，得到产品所属归类，将该类表示为用户特征，如下所示：

$$F_A(u)=\{c_1:w_1,c_2:w_2,\cdots,c_k:w_k\} \tag{8-11}$$

8.3.3　时间维度和空间维度上的用户特征建模

前面章节分别从两大类反馈特征、4小类用户数据来构建用户特征，因此可用如下表达式表达用户特征：

$$F(u)=F_D(u)\cup F_G(u)\cup F_T(u)\cup F_A(u) \tag{8-12}$$

然而，用户兴趣是随时间变化而变化的，特别是在今天这种信息过载的情况下，信息与物质丰富造成用户兴趣的转移越来越快，用户专注某一兴趣的时间越来越短，用户特征随时

间的变化越来越强。针对特定时间，期望使用相邻时间的用户特征来描述用户。因此，将用户反馈数据按时间段划分，每一段数据集中作为用户该段时间用户特征。假设距离目标时间的时间差为 t，可以利用正态分布计算权重：

$$w(t)=\mathcal{N}(u,t)=(\mu,\sigma^2) \tag{8-13}$$

$$\hat{F}(u)=\sum_{t=1}^{T}w(t)\times F(u,t) \tag{8-14}$$

用户数据分布具有不均匀性，存在长尾分布，即多数用户的数据比较稀疏。并且，由于加入了时间维度，用户数据将会变得更加稀疏，可能导致某段时间用户特征为空。在这里可以从空间上，利用目标用户好友数据或者目标用户相似用户特征来弥补数据稀疏问题。针对目标用户，可以利用显式特征来计算相似用户，将相似用户作为补充数据。

对于显式信息全局空间向量，0-1 型特征信息，比如性别，可以直接利用 0-1 结果计算相似度，如果性别相同，相似度为 1，否则为 0；对于长度型特征信息，比如年龄或地理位置，采用距离倒数作为相似度，两者距离越近，相似度越接近 1。其计算距离可以称为dis_1。对于显式信息中的关系型数据，可以采用两者在图中最短距离的倒数作为相似度，距离越近，相似性越大；也可以采用共同好友数量作为相似度，共同好友越多，其相似性越大。其计算距离可以称为dis_2。对于隐式信息中的行为信息，其信息表示为标签向量，可以采用扩展 Jaccard 或者 Cosine 计算距离，由于向量自身带有权重，其计算距离可以称为dis_3。对上述每一种距离赋予权重，融合计算结果，$\text{Dis}=\text{dis}_1+\text{dis}_2+\text{dis}_3$。

假设可以计算出目标用户相似用户，那么就可以构建二维正态分布。从空间维度（S）和时间维度（T）融合相似特征，作为用户最终的特征表示。采用二维正态分布计算其权重，其中权重计算公式为

$$w(S,T)=\frac{1}{2\pi\sigma_S\sigma_T\sqrt{1-\rho^2}}\mathrm{e}^{\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(S-\mu_S)^2}{\sigma_S^2}-2\rho\frac{(S-\mu_S)(T-\mu_T)}{\sigma_S\sigma_T}+\frac{(T-\mu_T)^2}{\sigma_T^2}\right]\right\}} \tag{8-15}$$

其中，μ_S为 0，表示空间距离为 0 时，相似度最大；μ_T为目标时间；σ_S和σ_T根据计算的扁平程度调整。为计算方便，令 $\rho=0$，即时间维度和空间维度不存在关联性。那么，对于目标用户，目标时间的用户特征模型为

$$\hat{F}(u,t)=\sum_{i=1}^{n}\sum_{j=1}^{T}w(S,T)\times F(u_i,t_j) \tag{8-16}$$

8.4　跨网络用户偏好可预测性分析

8.4.1　实验数据集与用户建模方法

为了验证用户特征建模有效性，利用采集的部分真实社交网络 Facebook 用户数据和消费网络 eBay 用户数据进行实验。由于用户特征建模是针对单个用户数据分析，因此在本节中，针对两个用户 lynne 和 lowell，进行用户特征建模作为本章方法的应用用例。这两个用户在 Facebook 和 eBay 中均有账户，数据较丰富且具有代表性。除了用户显式反馈外，还采集了以上用户在 2014 年 1 月～2014 年 6 月的隐式反馈数据，其数据量如表 8-1 所示。

表 8-1 用户数据量

	Facebook		eBay
	消息量	喜好量	交易量
lynne	8143	316	128
lowell	3955	2973	76

针对采集的用户数据，首先分 4 个维度对用户进行特征建模，特别是用户隐式文本数据，采用基于知识库理解的方法并给出比较策略；然后针对用户时间维度构建用户特征演化模型给出用户特征变化模型；最后，融合空间维度相似用户特征，证明融合相似特征可以得到更准确的用户特征。

对于人文特征信息，首先统计用户在 Facebook 和 eBay 中的特征空间向量，其中，Facebook 中共有 18 个维度，主要包括〈id，about，age_range，birthday，context，education，email，gender，hometown，languages，link，locale，location，name，quotes，relationship_status，religion，work〉；而 eBay 共有 6 个可用维度，分别是〈UserID，AboutMePage，Email，RegistrationAddress，RegistrationDate，Status〉。用户显式特征数据主要面临 3 个问题，数据泛化、数据融合和数据平滑，其中，数据泛化是指将稀疏分布的数据泛化成等级，比如用户年龄（age）、地理位置（location）等，将其泛化成年龄区间（age_range）、地理区间（location_range）；数据融合是指相似数据融合，比如年龄和生日，统一融合成年龄表示；而最重要的问题是要面对数据平滑问题，限于用户的公开性，用户数据不一定是完整的，在一些维度，用户数据可能是空缺的。对于一些数据，比如工作（work），如果不存在，就以空表示，而比如年龄，则利用其中位数表示。对于用户所在社区，扩展用户的好友，由于所选用户比较活跃，其好友数量较大，lynne 用户有好友 316 个，lowell 有好友 2973 个。

对于隐式文本数据，使用 Freebase 知识库来分析数据。Freebase 是一个由元数据组成的大型合作知识库，整合了网络资源和社区成员贡献的数据。截至 2015 年，Freebase 共收录 47 543 893 个主题，2 919 496 429 个实体。利用 Freebase 学习用户文本数据特征，将采用实体与概念标注一体化模型学习用户特征。本文采用三元文法，即

$$P(E,C)=\prod_{i=1}^{n}P(c_i|e_i)\times P(e_i|e_{i-1}e_{i-2}\cdots e_1)\approx\prod_{i=1}^{n}P(c_i|e_i)\times P(e_i|e_{i-1}e_{i-2}) \qquad (8\text{-}17)$$

其中，当 $i=1$ 时，取 $P(e_i|e_{i-1}e_{i-2})=P(e_1)$。当 $i=2$ 时，取 $P(e_i|e_{i-1}e_{i-2})=P(e_2|e_1)$。进一步地，还可以利用 LDA[25] 主题学习模型、CPK[26] 主题模型作为比较学习用户特征，对比分析其有效性。在本实验中，使用 F_1 值计算用户特征计算的准确度。

隐式行为反馈：在用户行为中，假设用户行为都是正情感的，即用户有点击行为，用户便偏好该类产品。使用基本的统计方法给出用户偏好，其中次数表明权重。如式(8-17)所示。

在基于 4 类数据建模的基础上，分别从时间维度和空间维度考虑用户特征。在时间上，特别是隐式反馈数据，将用户反馈按时间段划分（本例中按月份划分）分别计算用户每个月份的特征。然后利用 Jaccard 计算用户月份之间的距离。其计算公式如下所示：

$$J_u(t_i,t_j)=\frac{F_u(t_i)\cap F_u(t_j)}{F_u(t_i)\cup F_u(t_j)} \qquad (8\text{-}18)$$

其中，$F_u(t_i)$值用户在时间段t_i内用户 u 的特征。

在空间上，构建二维正态分布，为了简单计算，假定时间维度与空间维度不存在关联关系，即假定 $\rho=0$。对于目标时间 t 目标用户 u，其中，空间$\mu_S=0$，时间$\mu_T=t$。仅需要通过调整参数σ_S和σ_T，定义其区间为[0－10]和[0－10t]，其中条件距离为 0.1。同样，将单个用户隐式反馈数据按月划分，在本实验中，集中关注 4 月份的用户特征建模，因为可以从时间维度上补充。在评价标准中，使用 F_1 值计算用户特征学习的准确度。

8.4.2　实验结果与分析

对于文本数据理解，采用 LDA 和 CPK 方法作为对比，采用 F_1 作为评价指标。如图 8-5 所示。从实验数据中可以看出，TUMK 效果优于所有策略，而使用外部语料库的 CPK 方法平均要比 LDA 方法更优秀。LDA 作为一种非监督学习方法，只能从词和文档学习主题，受词频影响很大，而且无法理解词义的深层概念，所以往往得到一些词频比较大、但却没有意义的关键词。CPK 方法通过语料库学习，可以理解词内部含义，但是无法利用词序，破坏了词与词之间的关联关系，所以虽然能在一定程度上理解用户概念，但是无法准确学习。TUMK 不仅考虑词的深层含义，而且有效识别实体，平均比 LDA 准确 30%，在$k=15$ 时，效果达到最优，比 CPK 有效 17%。

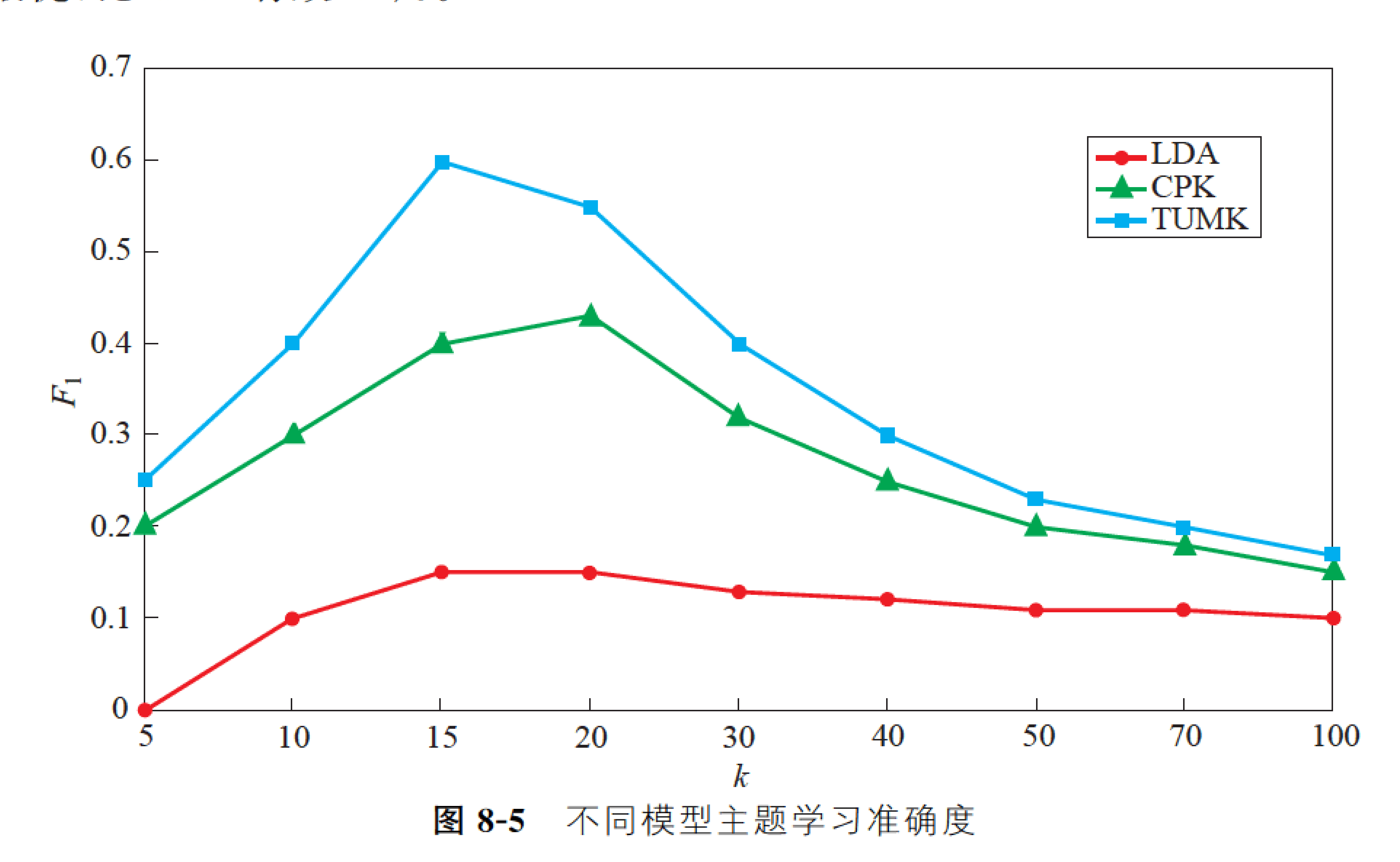


图 8-5　不同模型主题学习准确度

小结

本章针对用户在社交网络和消费网络中的显式反馈和隐式反馈数据进行特征建模，并对不同类型数据采用针对性的计算方法，最后结合真实数据给出实验证明。

首先，将用户特征划分为两大类：显式反馈和隐式反馈，其中，显式反馈又可划分为人文特征数据和社区特征数据，隐式反馈可以划分为文本类数据和行为类数据。针对不同的数据，采用不同的计算方法，比如人文特征数据，通过数据规整、数据泛化以及数据融合 3 个步骤，构建机器可理解的特征。针对文本类隐式反馈数据，明确以往解决方法所面临的问题，利用外部语料知识库提出基于统计的实体与概念标注一体化模型，用以准确学习用户特征。

其次，在基于用户数据学习的基础上，提出用户特征的时空模型。为了学习用户特征随时间转移的特征，加入时间维度，将用户数据按时间切片；为了解决数据稀疏问题，加入空间维度，将相似用户数据赋予目标用户。从时间和空间两个维度，构建为高斯分布，将相近特征加入目标用户特征。

最后，基于 TUMK，利用真实 Facebook 用户数据展开建模学习，并对实验结果进行分析。从如下角度分析实验结果：用户特征学习的准确性、用户特征时间维度演变性以及用户特征空间维度和时间维度的互补性。

本章基于用户不同类型数据构建用户特征，提出基于统计的实体与概念标注一体化模型，提出用户特征时空演变，并提出实验证明方法的准确性以及有效性。在下一章中，将利用构建的用户特征，预测用户在消费媒体中的偏好。

参考文献

[1] Yu Z, *et al*. TV program recommendation for multiple viewers based on user profile merging. User Modeling and User-Adapted Interaction 16.1. 2006. 63-82.

[2] Baltrunas L, Amatriain X. Towards time-dependant recommendation based on implicit feedback. In: Proc. of the Workshop on Context-Aware Recommender Systems (CARS'09). 2009.

[3] Fragoudis D, Likothanassis SD. User modeling in information discovery: an overview. In: Proc. of the Advanced Course on Artificial Intelligence. 1999: 17-43.

[4] Rich E. Users are individuals: Individualizing user models. Int'l Journal of Man-Machine Studies, 1983, 18(3): 199-214.

[5] Ardissono L, Torasso P. Dynamic user modeling in a Web store shell. ECAI. 2000.

[6] Ahmed A, *et al*. Scalable distributed inference of dynamic user interests for behavioral targeting. In: Proc. of the 17th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. ACM, 2011.

[7] Albanil P, *et al*. Improving the quality of services at the University of Oviedo. Sinergie Italian Journal of Management, 2013.

[8] Kobsa A. Generic user modeling systems. User Modeling and User-Adapted Interaction, 2001, 11(1-2): 49-63.

[9] Stonebraker M, Cetintemel U. "One size fits all": an idea whose time has come and gone. In: Proc. of the 21st Int'l Conf. on Data Engineering. IEEE, 2005.

[10] Lee H, Choi YS, Kim YJ. An adaptive user interface based on spatiotemporal structure learning. IEEE Communications Magazine, 2011, 49(6): 118-124.

[11] Jawaheer G, Szomszor M, Kostkova P. Comparison of implicit and explicit feedback from an online music recommendation service. In: Proc. of the 1st Int'l Workshop on Information Heterogeneity and Fusion in Recommender Systems. ACM, 2010: 47-51.

[12] White RW, Jose JM, Ruthven I. Comparing explicit and implicit feedback techniques for Web retrieval: Trec-10 interactive track report. In: Proc. of the 10th Text Retrieval Conf. (TREC-10). NIST, 2002.

[13] Lakiotaki K, Matsatsinis N, Tsoukiàs A . Multicriteria user modeling in recommender systems. IEEE Intelligent Systems, 2011, 26(2): 64-76.

[14] Iglesias JA. User modeling in changeable environments. In: Proc. of the 2012 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS). IEEE, 2012.

[15] Rossignol S, Pietquin O, Ianotto M. Training a BN-based user model for dialogue simulation with missing data. In: Proc. of the IJCNLP. 2011.

[16] Abel F, *et al*. Analyzing user modeling on Twitter for personalized news recommendations. In: Joseph A, *et al*. (eds.) User Modeling, Adaption and Personalization. Berlin, Heidelberg: Springer-Verlag, 2011. 1-12.

[17] Gao Q, Abel F, Houben GJ. GeniUS: generic user modeling library for the social semantic web. In: Jeff Z, *et al*. (eds.) The Semantic Web. Berlin, Heidelberg: Springer-Verlag, 2012. 160-175.

[18] Carmagnola F, *et al*. Tag-Based user modeling for social multi-device adaptive guides. User Modeling and User-Adapted Interaction, 2008, 18(5): 497-538.

[19] Liu Q, *et al*. Predicting user likes in online media based on conceptualized social network profiles. Web Technologies and Applications. Springer, 2014. 82-92.

[20] Schwarzkopf E, *et al*. Mining the structure of tag spaces for user modeling. In: 11th Int'l Conf. on User Modeling, 2007.

[21] Razmerita L, Angehrn A, Maedche A. Ontology-based user modeling for knowledge management systems. User Modeling 2003. Berlin, Heidelberg: Springer-Verlag, 2003. 213-217.

[22] Skillen KL, *et al*. Ontological user profile modeling for context-aware application personalization. Ubiquitous Computing and Ambient Intelligence. Berlin, Heidelberg: Springer-Verlag, 2012. 261-268.

[23] Shepstone SE, Tan ZH, Jensen SH. Demographic recommendation by means of group profile elicitation using speaker age and gender recognition. In: INTERSPEECH. 2013. 2827-2831.

[24] Kau AK, Tang YE, Ghose S. Typology of online shoppers. Journal of Consumer Marketing, 2003, 20(2): 139-156.

[25] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. Journal of Machine Learning Research, 2003, 3: 993-1022.

[26] Song Y, Wang H, Wang Z, *et al*. Short text conceptualization using a probabilistic knowledgebase. In: Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence, Vol. 3, AAAI Press, 2011. 2330-2336.

第 9 章 社交网络用户在强关联消费网络中的偏好预测

所谓融合异质社交网络信息的预测方法，是试图利用用户在社交网络上的数据预测用户在消费媒体中的偏好。融合多源数据的预测方法，其基本思想是将一个数据源学习的用户或对象特征或知识应用到另外一个数据源中，综合利用多数据源的知识共同解决数据稀疏问题，从而提高预测算法的准确率。特别地，当出现一些用户在某些网络中个性化特征较少、数据稀疏性大(比如在消费媒体中)，而在另一个网络中行为反馈信息较多(比如社交网络中)时，通过整合用户在社交网络中的用户特征可以建立一种融合异质网络信息的预测模型，使多个用户特征相互补充。如果用户在社交网络与消费媒体之间的账户存在显式关联，例如人名、电话号码、电子邮件等信息完全一致等，则称用户账户强关联。

在本章的最后，基于前述用户特征建模基础以及强关联的账户绑定，提出一种融合异质社交网络信息的消费媒体用户偏好个性化预测模型 FHSP(fusion heterogeneous social network information and online shopping for user preference prediction)。

9.1 概述

如图 9-1 所示，账户强关联的个性化预测分为如下几个步骤：首先利用第 8 章给出的用户特征建模模型，学习用户在社交网络和消费网络中的模型，学习产品特征；然后利用特定预测方法融合社交网络特征和消费媒体特征，预测用户在消费媒体中的偏好。预测训练模型可以包含反馈信息，通过不断反馈学习优化模型参数。

形式化的预测过程可如下表示：假设给定一个社交网络G^S和一个目标媒体G^Y，并假设社交网络中用户均拥有较详细的个人信息，而用户在消费媒体中没有足够的个人信息和历史行为。社交网络$G^S=(U,F^S)$包含用户个人信息，$U=\{u_1,u_2,\cdots,u_n\}$用于表示用户，并且

$F^S = D^S \cup T^S$表示用户在社交网络中的特征。其中,$D^S = \{d_1, d_2, \cdots, d_k\}$代表用户的人文特征信息,比如性别、年龄等;而$T^S$代表用户的历史行为特征,通常由用户发表的信息得到。同样地,用户的消费媒体特征可以表示为$G^Y = (U, C^Y)$,其中包含相同的用户U,而$C^Y = \{c_1, c_2, \cdots, c_m\}$表示用户在消费媒体中的偏好。

异质网络预测问题:给定社交网络G^S和消费媒体G^Y,利用用户在社交网络中的人文特征和历史行为信息,能否有效预测用户在消费媒体中的偏好?

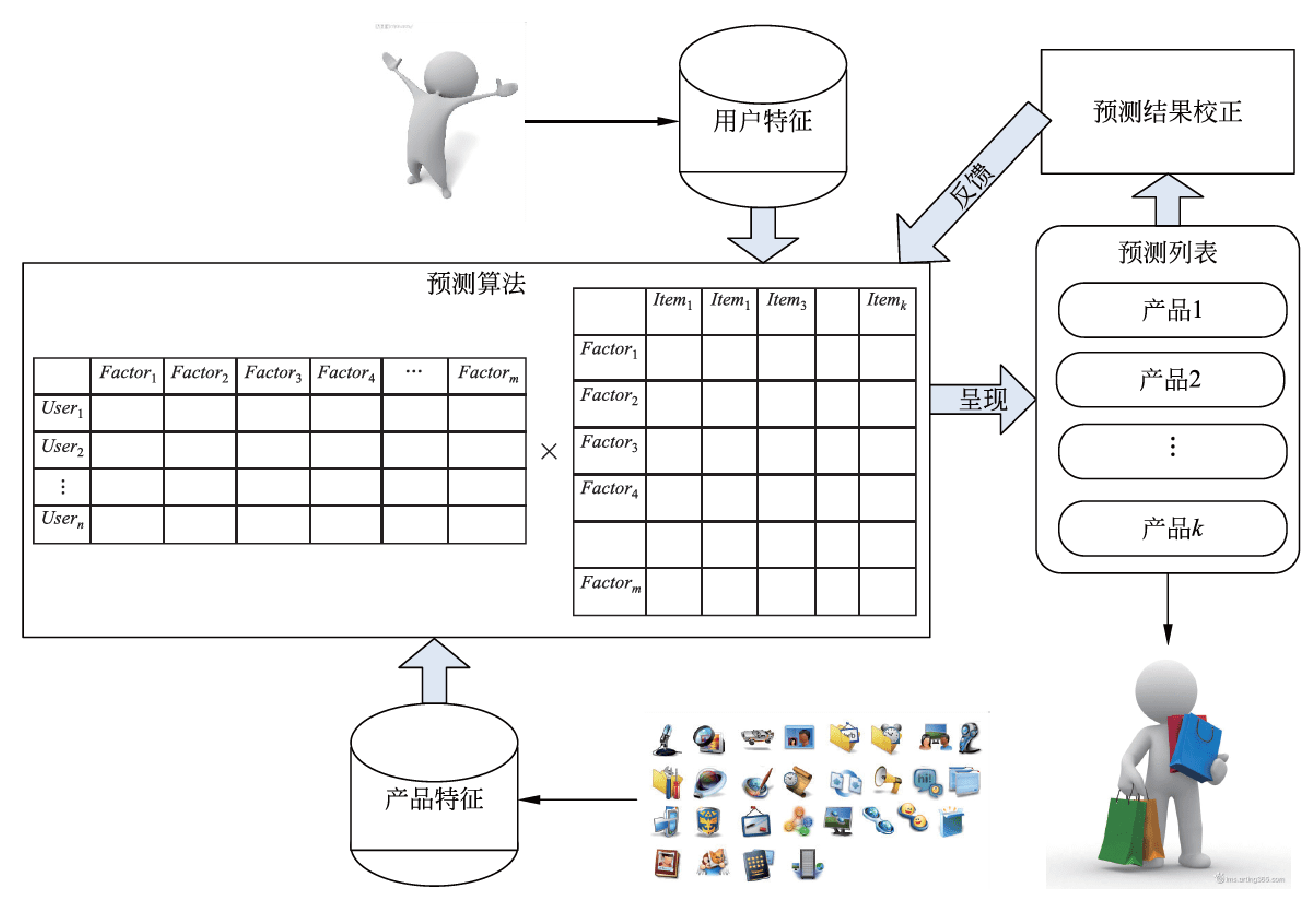


图 9-1　融合异质社交网络数据的消费媒体个性化预测流程图

9.2　相关研究工作

在大数据互联网环境下,如果用户不能及时、有效地获得自己感兴趣的消息,就会导致信息过载问题。个性化服务是信息爆炸的必然结果,不同于普通的搜索服务,个性化服务具有更高的服务质量。因此,推荐技术被深入学习并且广泛应用于互联网社交、消费领域等。如图 9-2 所示,可以从 3 个角度研究预测相关工作,分别是预测模型、预测方法以及评价指标。

9.2.1　个性化预测技术

目前学术界和工业界有几种广泛应用的个性化预测技术,包括基于内容的模型(content-based filtering)[1,2]、基于协同过滤的模型(collaborative filtering)[3,4]、基于案例的特殊模型(case study)[5-7]以及融合多资源的预测模型(hybrid-based model)[8,9]。

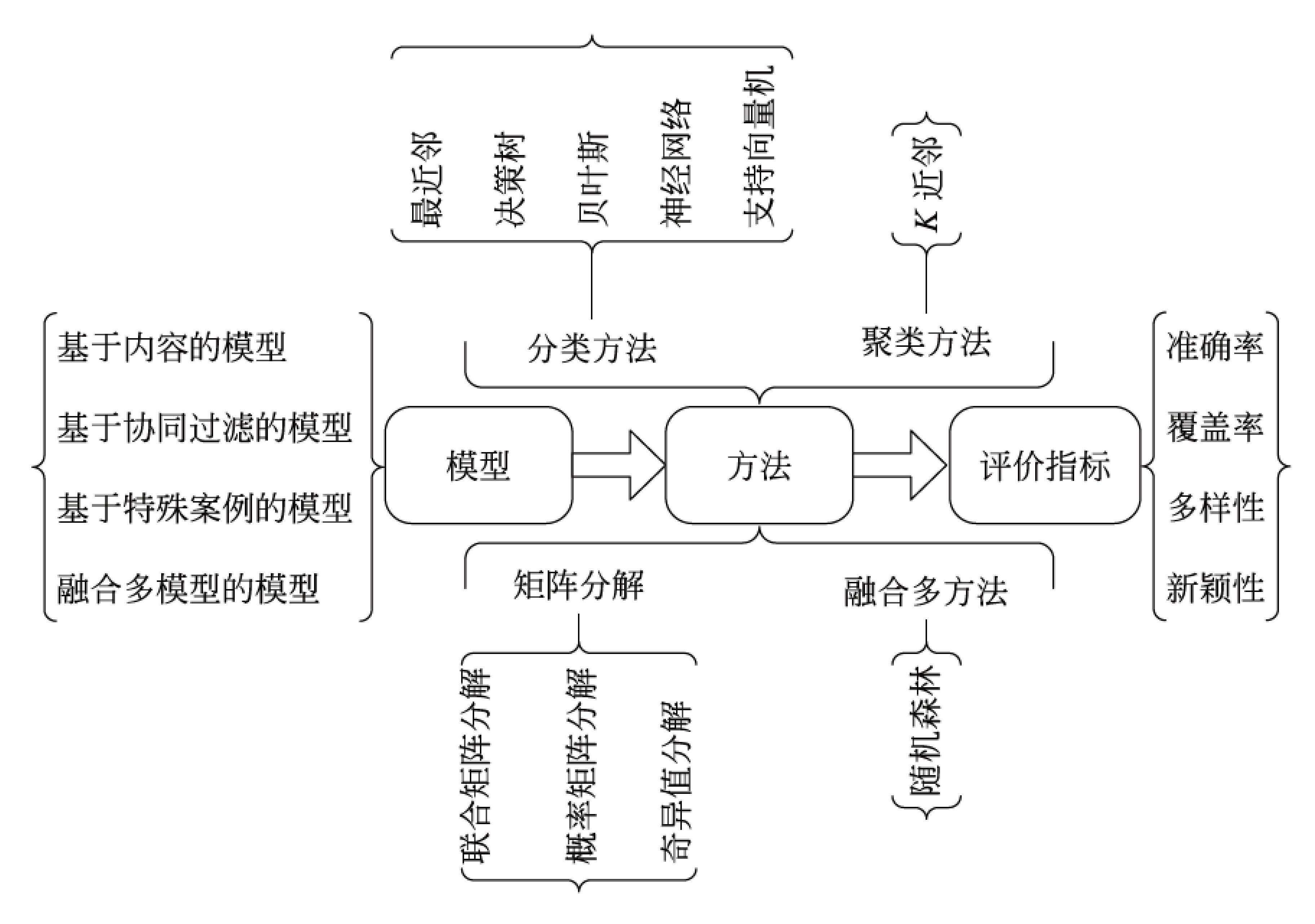


图 9-2　个性化预测的组成部分

基于内容过滤预测是指根据用户历史记录中喜欢的物品，向其推荐内容上相似的其他物品，其模型主要关注产品内容的相似度。该模型通常首先统计用户历史偏好记录，对记录中产品特征信息进行挖掘，得到能描述用户兴趣的特征信息。预测系统依据这些特征信息，向用户推荐相似产品。

如图 9-3 所示，基于内容的预测模型通常可以分为 3 部分：产品解析（content analyzer）、用户特征学习（profile learner）和过滤选择（filtering component）[10]。产品解析针对产品特征信息的无结构性，比如使用文本描述产品，通过一个预处理过程将无结构的产

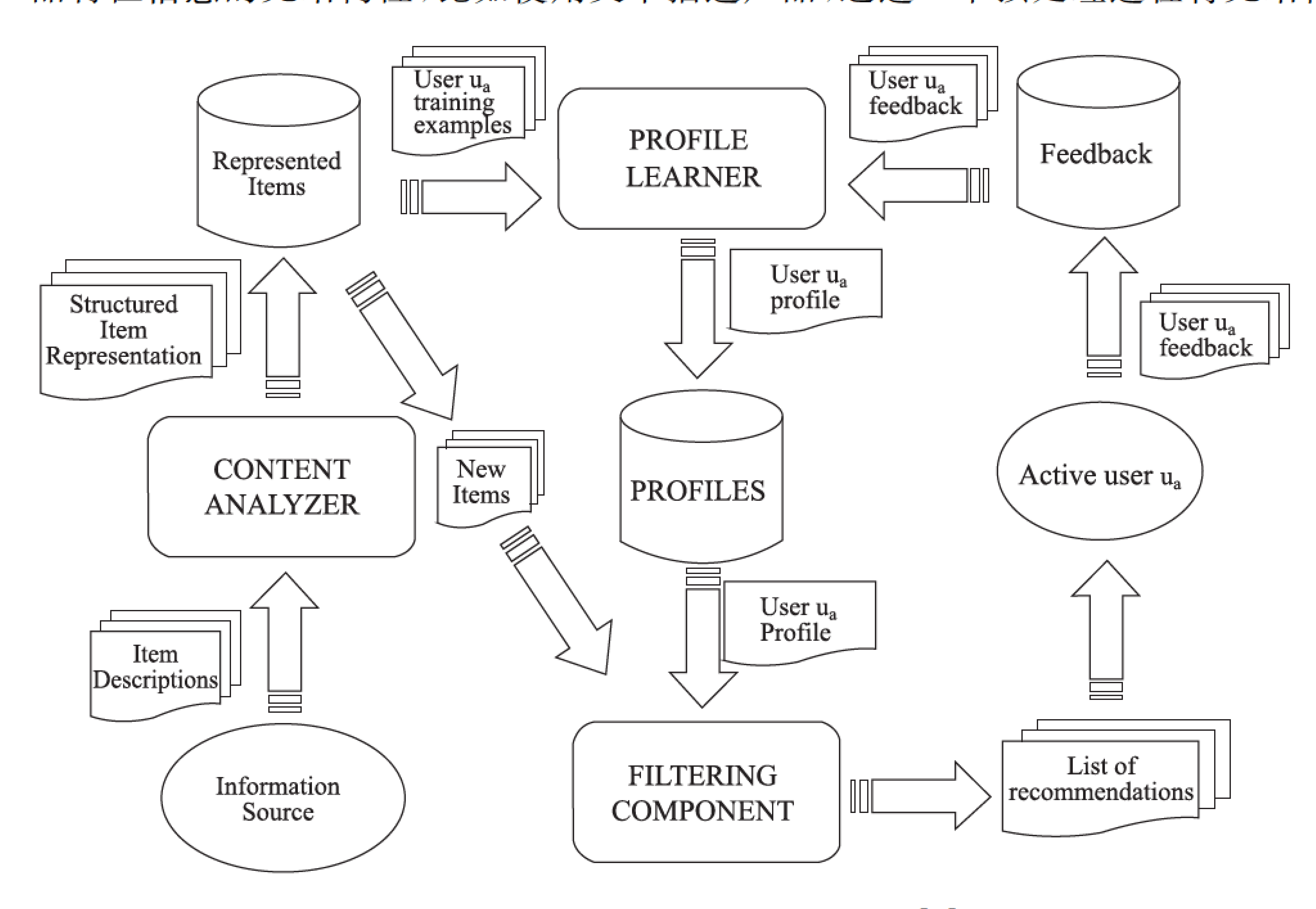


图 9-3　基于内容的个性化预测流程图[11]

品信息解析成结构化信息,作为用户特征学习和过滤选择的数据输入。用户特征学习通过收集能代表用户偏好的数据构建用户特征,通常该过程通过用户的历史记录,利用机器学习方法推测出用户偏好。过滤则是匹配产品特征和用户偏好特征的一个过程。下面给出了基于内容过滤的推荐的计算方式,其中,f_i 表示产品的一种特征,sim 表示匹配产品特征和用户特征的机器学习方法。

$$C(c)=\{f_1,f_2,\cdots,f_k\} \tag{9-1}$$

$$U(u)=\frac{1}{N(u)}\sum_{i\in N(u)}C(i) \tag{9-2}$$

$$p(u,c)=sim(U(u),C(c)) \tag{9-3}$$

基于内容的预测可以保持用户特征独立性(user independence)[11],不同于协同过滤依赖相似用户进行数据计算,基于内容的预测仅从用户自身的历史记录中计算用户特征,可以保持用户特征的个性化;计算显式透明(transparency):基于内容的预测利用喜好产品的特质,计算方法容易理解;不存在新产品冷启动问题:由于基于内容的预测利用的是产品的特质而非用户的特征,所以不会产生产品冷启动问题(item cold start)。但是,基于内容推荐可能没有足够的产品特征信息可供分析(limit content analysis),多数情况下,产品的特征分类比较稀疏,可能需要外部语料库协助理解;存在信息封闭情况(over specialization):基于内容的预测只能给出用户历史喜好的产品,不能跳出用户兴趣范围向用户推荐一些“意想不到”的产品;存在用户冷启动问题(user cold start):如果用户没有历史记录,那么基于内容的预测将因没有输入数据而无法预测用户偏好。

不同于基于内容的预测模型,基于协同过滤的预测模型依赖系统中其他用户对产品的评价,其假设是相似用户通常喜好相似产品。协同过滤可以解决一些基于内容预测的弊端,比如协同过滤是基于系统中相似用户的评价,而不是依赖于产品本身信息,多数情况下产品信息准确度很低;一些没有基本信息的产品,根据其他用户的历史行为,依然可以推荐给系统中的用户;而且协同过滤会跳出用户已有的兴趣范围,向用户推荐一些“意想不到”的产品。

基于协同过滤的推荐可以分成两大类:基于邻居的协同过滤(neighborhood based)和基于模型的协同过滤(model based)。基于邻居的协同过滤又可以分为两类:基于用户的协同过滤(user based)和基于产品的协同过滤(item based)。基于用户的协同过滤,比如GroupLens、Bellcore 以及 Ringo,直接利用邻居对产品的评价计算用户对一项产品的偏好程度。其他方法中有的会利用好友关系确定用户的邻居关系,有的利用用户资料相似程度计算用户邻居,有的则直接使用用户共同历史行为记录计算邻居。

$$\text{Jaccard}(w_{uv})=\frac{|N(u)\cap N(v)|}{|N(u)\cup N(v)|} \tag{9-4}$$

$$\text{Cosine}(w_{uv})=\frac{|N(u)\cap N(v)|}{|N(u)||N(v)|} \tag{9-5}$$

$$p(u,i)=\sum_{v\in S(u,K)\cap N(i)}w_{uv}r_{vi} \tag{9-6}$$

如上述公式所示:给定用户 u 和用户 v,令 $N(u)$表示用户 u 感兴趣的物品集合,$N(v)$表示用户 v 感兴趣的物品集合,可以通过 Jaccard 公式或余弦相似度计算 u 和 v 的兴趣相似

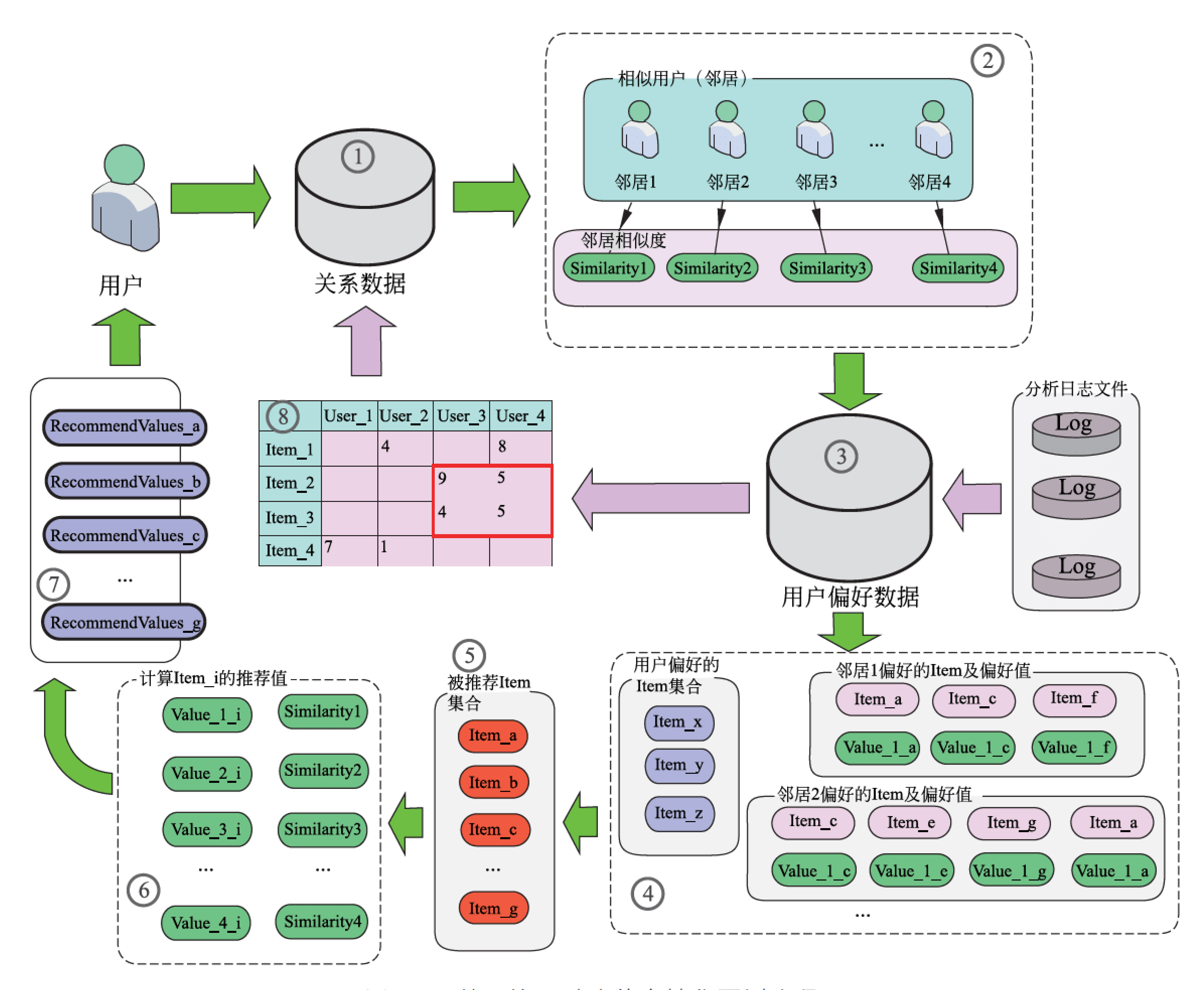


图 9-4　基于协同过滤的个性化预测流程图

度。$S(i,K)$表示和产品 i 最相似的 K 类产品的集合，$N(u)$表示用户 u 喜欢的物品的集合，w_{ij} 表示物品 i 和物品 j 的相似度，r_{ui} 表示用户 u 对物品 i 的感兴趣程度。

另一方面，基于产品的协同过滤主要计算用户是否对目标产品的相似产品有兴趣。这种方法中，两类产品是否相似取决于系统中是否有多数用户对两类产品行为或评价一致。给定物品 i 和物品 j，令 $N(i)$表示对物品 i 感兴趣的用户集合，$N(j)$表示对物品 j 感兴趣的用户集合，可用如下公式计算物品 i 和物品 j 的相似度：

$$w_{ij}=\frac{|N(i)\cap N(j)|}{\sqrt{|N(i)||N(j)|}} \tag{9-7}$$

在得到物品之间的相似度后，基于产品的系统过滤模型会给用户推荐和他感兴趣的物品前 K 个最相似的物品。可用如下公式计算用户 u 对物品 i 的感兴趣程度：

$$p(u,i)=\sum_{j\in N(u)\cap S(i,K)} w_{ij}\, r_{uj} \tag{9-8}$$

其中，$S(i,K)$表示和物品 i 最相似的 K 个物品的集合，$N(u)$表示用户 u 喜欢的物品的集合，w_{ij} 表示物品 i 和物品 j 的相似度，r_{ui} 表示用户 u 对物品 i 的感兴趣程度。

不同于直接使用系统中用户对产品的评价，基于模型的协同过滤首先利用这些评价关系学习预测模型，再利用模型判断用户对目标产品的兴趣度。这类模型中，主要思想是试图

学习用户与产品之间的特征关联关系。这些特征,对用户来说可以是用户的偏好特征,对产品来说可以是产品的分类特征。模型是训练两者之间的关联关系。这种学习过程能够挖掘出潜在的隐藏特征,代表性方法如基于贝叶斯的距离方法(Bayesian clustering)[12]、隐语义分析(latent semantic analysis)[13]、LDA 方法(latent Dirichlet allocation)[14]、最大熵(maximum entropy)[15]、Boltzmann Machines[16]、支持向量机(support vector machine)[17]和奇异值分解(singular value decomposition)[18]等。比如在利用隐语义模型的学习方法中:

$$p(u,i)=r_{ui}=p_u^{\mathrm{T}}q_i=\sum_{k=1}^{K}p_{u,k}q_{i,k} \tag{9-9}$$

其中,$p_{u,k}$表示用户 u 和第 k 类隐语义的关系,$q_{i,k}$表示物品 i 和第 k 类隐语义的关系。可以通过优化以下的损失函数来找到最合适的参数 p 和 q:

$$\begin{aligned}\min C&=\sum_{(u,i)\in K}(r_{ui}-\hat{r}_{ui})^2\\&=\sum_{(u,i)\in K}\left(r_{ui}-\sum_{f=1}^{F}p_{u,k}q_{i,k}\right)^2+\lambda\|p_u\|^2+\lambda\|q_i\|^2\end{aligned} \tag{9-10}$$

$\lambda\|p_u\|^2+\lambda\|q_i\|^2$是用来防止过拟合的正则化项,可以利用随机梯度下降法求解。它首先通过求参数的偏导数找到最速下降方向,然后通过迭代法不断地优化参数。

$$\frac{\partial C}{\partial p_{uk}}=-2q_{ik}+2\lambda p_{uk} \tag{9-11}$$

$$\frac{\partial C}{\partial q_{ik}}=-2p_{uk}+2\lambda q_{ik} \tag{9-12}$$

$$p_{uk}=p_{uk}+\alpha(q_{ik}-\lambda p_{uk}) \tag{9-13}$$

$$q_{ik}=q_{ik}+\alpha(p_{uk}-\lambda q_{ik}) \tag{9-14}$$

除了基于内容和基于协同过滤的推荐模型,在一些特殊背景或特殊数据下,很多学者依据其具体背景提出很多优秀的基于特殊案例的预测模型。其中大致可以分为 3 类:基于人文特征(demographical based)[5]、基于领域知识(knowledge based)[6]以及基于社区(community based)[7]的模型。基于人文特征的方法是利用用户的一些人文特征,比如性别、年龄、地理位置等,向用户推荐相关产品。很多在线系统利用用户的人文特征进行预测,特别是在用户数据稀疏、特征差异与行为差异关联性比较大的背景下。比如,用户登录一个站点时,其语言依据用户所在国家而判断。基于领域知识的模型是利用特定领域知识判断用户对产品的需求性。最著名的模型是基于案例(case based)的推荐模型,在这种系统中,通过计算用户需求与推荐结果之间的相似程度预估推荐产品。另一种基于领域知识的模型称作基于限制(constraint based)的推荐模型。类似于基于案例的模型,两者都是使用领域知识,收集用户需求,给出推荐结果与相关解释。两者主要区别在于计算方式不同,基于案例的模型利用需求与结果相似度给出推荐结果,而基于限制的模型是利用提前构建的知识框架和显式规则给出推荐结果。基于领域知识的模型在系统刚启动、用户与产品数据稀疏的情况下,效果较好,但是随着新数据的增多及新知识的增长,模型效果逐渐衰弱。基于社区的模型通常利用用户的一些好友关系,向用户推荐好友偏好的产品。大量证据显示,用户倾向于依赖好友的推荐。随着社交网络的发展,用户好友关系或所在社区的显式性,这种方法逐渐被重视和使用。文献[7]证明,利用用户好友关系可以有效地促进预测结果的准确度。

以上模型都存在一些优势和劣势，如果将这些模型融合，发挥优势，避免缺陷，将能够得到更好的结果，这种策略通常称为融合多资源的模型。Chen[8]提出融合协同过滤模型和子序列匹配算法，可以发现用户和产品之间的潜在关联关系，优化用户学习过程。除了融合多模型外，近几年，很多学者利用媒体外的数据，比如相关联的社交网络数据，做推荐预测。该类模型也称为融合多资源的推荐模型。Liu[19]融合异质社交网络用户自我描述对消费媒体行为进行预测，证明其简短的个人描述也是非常有效的特征。融合多资源的模型可以有效地解决一些模型所面临的问题，但融合多资源模型需要耗费更多的计算资源并需要更多的用户数据。

9.2.2　个性化预测方法

上一节详细描述了推荐方法相关的模型，本节基于以上模型对典型的预测算法展开讨论。预测算法可分为 4 类：基于分类的算法（classification）[20-23]、基于聚类的算法（clustering）[24]、基于矩阵分解的算法（matrix factorization）[25-27]以及融合多方法（hybrid）的算法。

一个分类器依据能代表用户或产品的特征或标签空间，将用户或产品划分到不同类别中。例如，一部电影可以根据主角亲吻次数和打斗次数，将电影划分为爱情或动作类。在分类算法中，典型的算法包括最近邻居算法（KNN）[20]、决策树算法（DT）[21]、贝叶斯算法（NB）[22]、神经网络算法（ANN）[23]以及支持向量机算法（SVM）[17]。最近邻居算法试图将与目标相似的邻居属性直接赋予目标，比如想知道一个目标 q 所属的类别，已知集合 $X=\{\{x_1,l_1\},\{x_2,l_2\},\cdots,\{x_n,l_n\}\}$，其中，$x_j$是第 j 个种类，l_j是 x_j所属的类别，那么计算最相似的 k 个邻居子集 $Y=\{\{x_1,l_1\},\{x_2,l_2\},\cdots,\{x_k,l_k\}\}$，其中，$Y\subset X$，并且 $\sum_{i}^{k}\mathrm{dis}(q,x_i)$ 最小。那么，统计 Y 的所有节点属性，将属性赋予 q，则 q 的属性为 $L_q=\sum_{i}^{k}l_i$。

决策树算法试图构建一种分类树，其中每个节点是一个属性，叶子节点是类别。比较著名的分类树有 CART、ID3、C4.5、SLIQ、SPRINT，这些决策树在计算分支的算法不同，比如 CART 利用最大信息收益，其计算公式如下：

$$\Delta_i=I(\text{parent})-\sum_{j=1}^{k_i}\frac{N(v_j)\times I(v_j)}{N}\tag{9-15}$$

其中，k_i是特征 i 的值；N 是总的训练数目；j 是利用第 j 个特征进行分支；I 用于计算纯净率。决策树将产品按特征进行划分，而没考虑特征之间的关系。

朴素贝叶斯方法基于条件概率模型关联特征之间的关系，比如每一个产品有 N 种属性（$A_1,A_2,\cdots,A_N$），那么预测某产品 C 所属类别，就是发现 C 的最大后验概率，其中，朴素贝叶斯假设特征之间具有条件独立性：

$$P(C_k\mid A_1,A_2,\cdots,A_N)\propto P(A_1,A_2,\cdots,A_N\mid C_k)P(C_k)\tag{9-16}$$

不同于朴素贝叶斯的单链节点，神经网络模仿神经元结构，通过调整节点之间相互连接的关系达到处理信息的目的。为了解决低数据维度无法进行有效分类的问题，有学者提出支持向量机模型，通过提高维度计算最大分隔距离，给出最优的分类算法。

$$w\cdot x+b=0\tag{9-17}$$

$$\text{Margin} = \frac{2}{\| w \|^2} \tag{9-18}$$

$$f(x)=\begin{cases}1, & w \cdot x+b \leqslant -1 \\ -1, & w \cdot x+b \leqslant -1\end{cases} \tag{9-19}$$

分类方法通常也称作有监督学习方法，通过记录学习模型。相反地，聚类可以看作是一种无监督学习方法，通常试图将相似的产品放到同一类别中，学习类别之间的不同。其中比较典型的聚类算法是 k-means[24]。k-means 试图将 N 个产品聚类到 k 个分类中，使得每个类别中节点距离自己类别中心最短，最终计算最短距离即可得到最优聚类。

$$E=\sum_{1}^{k} \sum_{n \in S_j} d(x_n, \lambda_j) \tag{9-20}$$

k-means 算法通常需要一些预知识以计算节点之间的距离，而且聚类受起始选择点的影响很大，并且受长尾效应很大，可能会产生空节组；相关改进算法包括 Density-based clustering、Hierarchical clustering 等。

近几年，特别是受 netfix 推荐算法大赛的影响，越来越多的学者将注意力转向矩阵分解算法。矩阵分解目标就是把用户-项目评分矩阵 R 分解成用户因子矩阵和项目因子矩阵乘的形式，即 $R=UV$，这里，R 是 $n\times m$，U 是 $n\times k$，V 是 $k\times m$。直观地表示如图 9-5 所示。

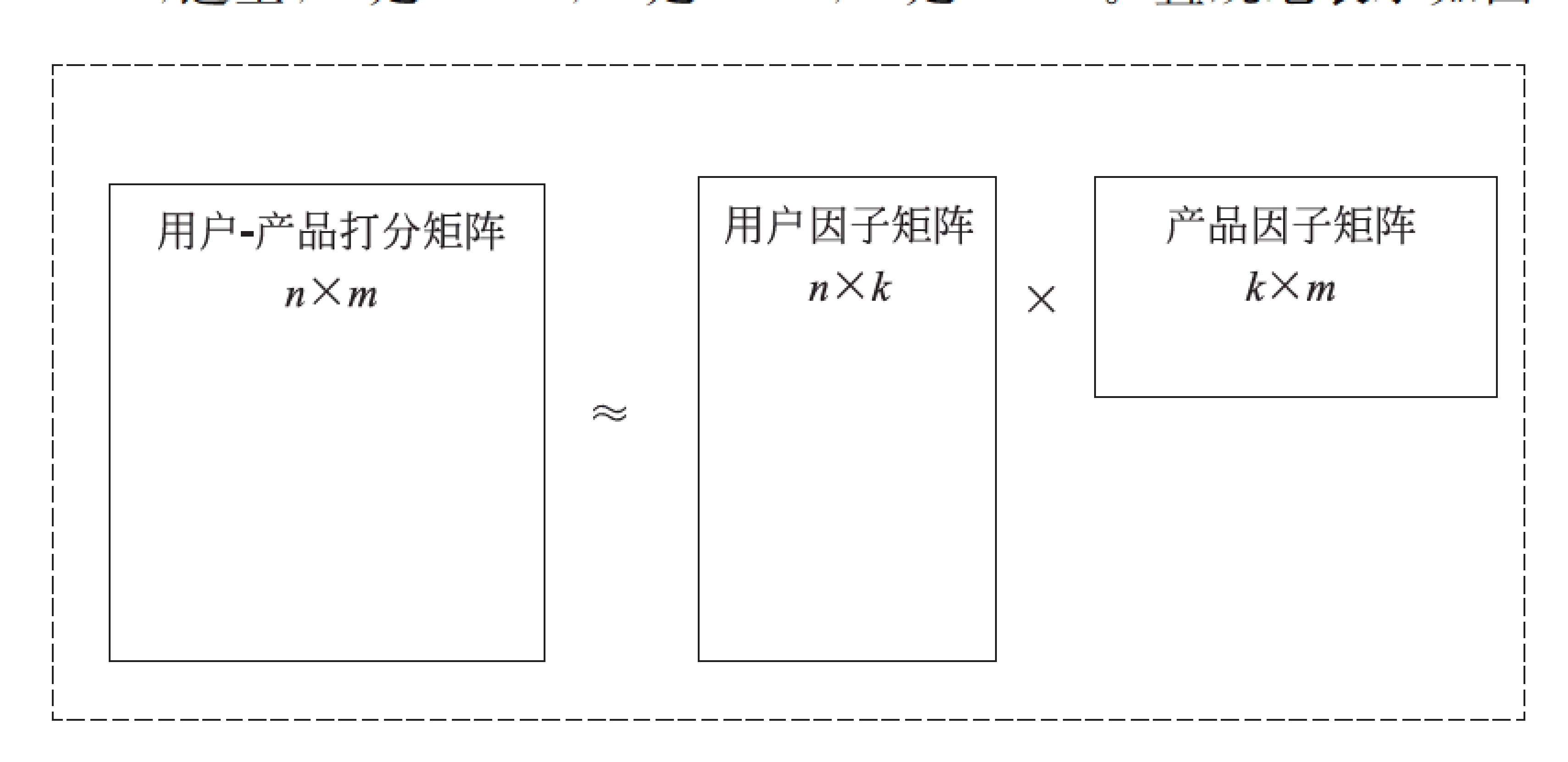


图 9-5　矩阵分解

朴素矩阵分解会产生负值，这些值无法解释现实的问题，为了有效处理大规模数据，文献[25]提出使用非负矩阵分解(non-negative matrix factorization)以解决预测排名问题。非负矩阵相较于朴素矩阵分解，具有实现上的简便性、分解形式和分解结果可解释性以及占用存储空间少等优点。用户和产品特征由于打分的随机性与用户针对性，数据会产生偏差，文献[26]从概率的角度预测用户的评分，假设用户和商品的特征向量矩阵都符合高斯分布。由于数据的稀疏性，基于迁移学习的想法，通过一个较为稠密的用户产品打分矩阵投射到一个子空间坐标系中，可以缓解数据稀疏性问题。文献[27]提出协同矩阵分解模型(collective matrix factorization，简称 CMF)，CMF 可以同时分解多个网络的打分矩阵，避免单数据源的数据稀疏问题。

9.2.3　推荐系统评价指标

推荐系统常用的评测推荐效果的方法包括用户调查(user study)、离线实验(offline experiment)和在线实验(online experiment)。用户调查是指让一些实际用户在真实的推荐

系统中进行操作,并根据用户的行为和答案熟悉系统的效果。离线实验利用用户历史日志记录构建离线数据集,将数据集分割为训练集和测试集,在训练集上训练推荐模型,在测试集上进行预测,并通过指定的离线指标评测算法在测试集上的预测结果。在完成离线实验和用户调查后,将推荐系统上线做 A/B 测试,与原有的算法进行比较。A/B 测试是一种在线评测算法的实验方法,它通过一定的规则将用户随机分成几组,并对不同组的用户采用不同的算法,然后通过统计不同组用户不同的评测指标来比较不同算法。

推荐系统常用的评测指标包括以下几种:用户满意度、预测准确率、覆盖率、多样性、新颖性和惊喜度。其中,准确率用于度量一个推荐系统预测用户行为的能力,通常用于预测方法的评价。常用的指标包括评分预测和 Top N 推荐,评分预测是指预测用户对物品的评分值,一般通过 RMSE 或 MAE 计算:

$$\text{RMSE} = \sqrt{\frac{\sum_{u,i \in T} (r_{ui} - \hat{r}_{ui})^2}{|T|}} \tag{9-21}$$

$$\text{MAE} = \frac{\sum_{u,i \in T} |r_{ui} - \hat{r}_{ui}|}{|T|} \tag{9-22}$$

而 Top N 推荐,一般通过准确率/召回率或信息检索领域中常用的 NDCG 或 MAP 来评价。

$$P = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \tag{9-23}$$

$$R = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \tag{9-24}$$

$$\text{NDCG} = \frac{1}{N_r} \sum_{i=1}^{N} \frac{2^{l(i)} - 1}{\log(1+i)} \tag{9-25}$$

$$\text{MAP} = \sum_{N(i) \in \text{Positive}} \frac{\text{count}(i)}{i} \tag{9-26}$$

覆盖率描述一个推荐系统对长尾分布的发掘能力,是推荐系统能够推荐出来的物品占总物品集合的比例,若能够覆盖绝大多数用户的偏好,则说明该推荐系统具有多样性。新颖性是指给用户推荐那些他们以前没有听说过的物品,可以简单地用推荐结果的平均流行度来度量,若平均流行度高,则新颖性低,反之,则新颖性高。虽然推荐结果和用户的历史兴趣不相似,但却让用户觉得满意,那么可以说这个推荐结果的惊喜度较高。它和新颖性是有区别的,推荐的新颖性仅仅取决于用户是否听说过这个推荐结果。

$$\text{Coverage} = \frac{\left| \bigcup_{u \in U} R(u) \right|}{|I|} \tag{9-27}$$

$$\text{Diversity}(R(u)) = 1 - \frac{\sum_{i,j \in R(u), i \neq j} s(i,j)}{\frac{1}{2} |R(u)| (|R(u)| - 1)} \tag{9-28}$$

$$\text{Diversity} = \frac{1}{|U|} \sum_{u \in U} \text{Diversity}(R(u)) \tag{9-29}$$

9.3　强关联预测模型与方法

9.3.1　基于显式反馈的用户偏好预测

用户显式反馈是指通过直接询问得到的特征和信息，以及允许用户自己定义和修改的特征信息，比如用户年龄、性别、地理位置等信息。社交网络的属性决定了其上的用户显式反馈信息相对较多，因此可以将这些特征作为一个关键因素计算用户偏好。如第 8 章所述，针对用户显式反馈，首先统计用户所有可能的特征维度，建立一个维度空间 d，将这些用户的人文特征信息转变成特征向量，用户u_i特征可以表示为$x_i \in R^d$。

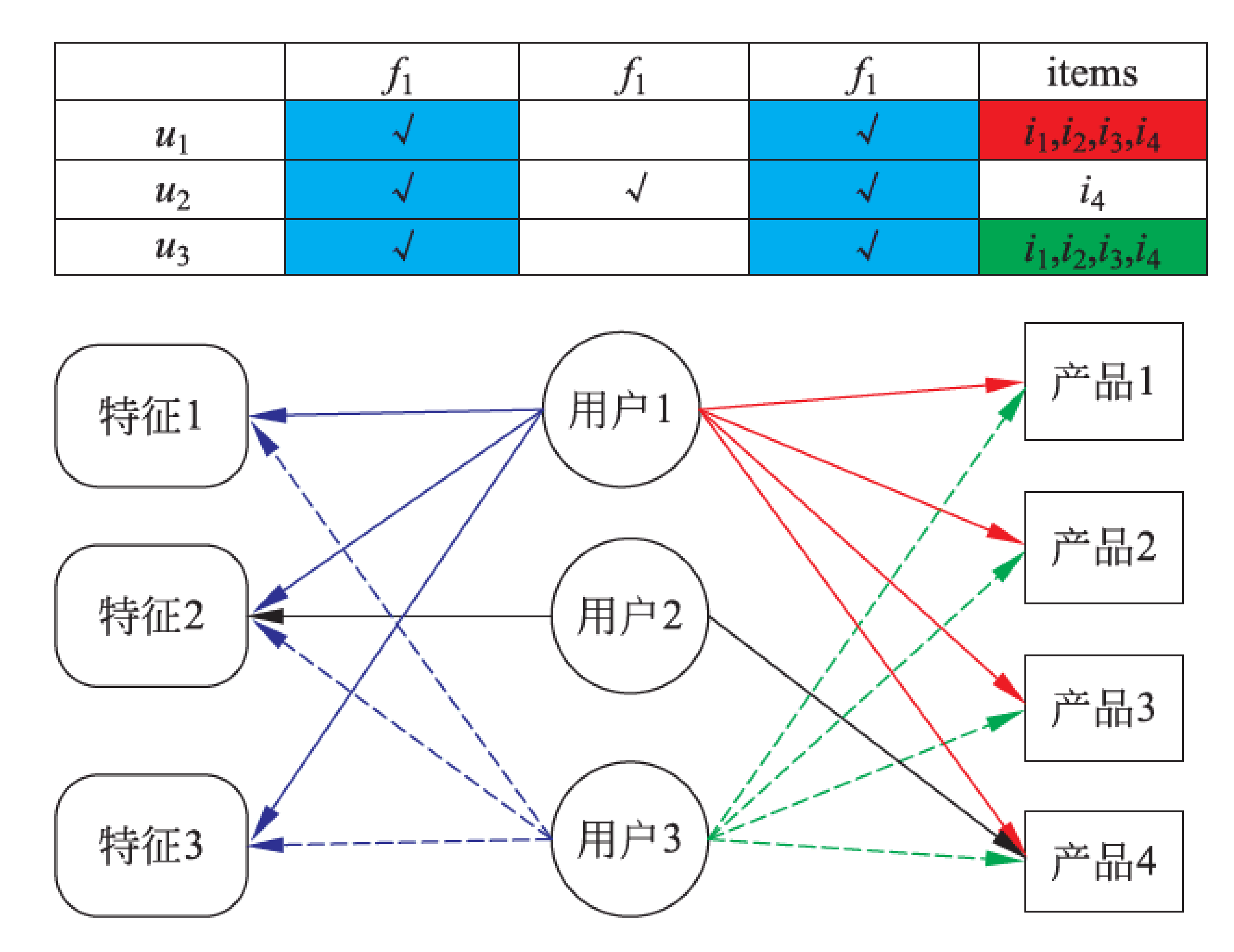
针对用户显式反馈特征，采用基于用户的协同过滤算法，如图 9-6 所示。即首先计算用户相似度，将相似用户偏好属性赋予目标用户。通常，最简单的用户相似计算方法如下所示：

$$\mathrm{Sim}(u_i,u_j)=\frac{{x_i}^{\mathrm{T}}x_j}{\sqrt{{x_i}^{\mathrm{T}}x_i}\sqrt{{x_j}^{\mathrm{T}}x_j}} \tag{9-30}$$

其中，x_i是用户u_i的特征，而x_j是用户u_j的特征。预测一个用户偏好消费媒体中一个类别的程度，可以利用上述相似度计算如下：

$$p_D(c_i\,|\,u)=\sum_{k=1}^{n}\mathrm{Sim}(u,u_k)\frac{\mathrm{like}(u_k,c_i)}{\sum_{c\in C}\mathrm{like}(u_k,c)} \tag{9-31}$$

其中，$\mathrm{like}(u,c)$表示用户 u 偏好类别 c 的程度。C 表示所有的类别。例如，用户喜欢 Music 类视频 3 个，喜欢 Sports 视频 2 个。那么可以得到$\frac{\mathrm{like}(u,\mathrm{Music})}{\sum_{c\in C}\mathrm{like}(u,c)}=0.6$和$\frac{\mathrm{like}(u,\mathrm{Sports})}{\sum_{c\in C}\mathrm{like}(u_k,c)}=0.4$。



	f_1	f_1	f_1	items
u_1	√		√	i_1,i_2,i_3,i_4
u_2	√	√	√	i_4
u_3	√		√	i_1,i_2,i_3,i_4

图 9-6　基于用户显式反馈特征的协同过滤算法

9.3.2　基于隐式反馈的用户偏好预测

社交网络用户通常会发布大量的个人信息，并与好友交流或者分享个人兴趣，如

Facebook 每天的消息量超过 120 亿条。用户发布的消息往往包含丰富的信息，并且具有实时性。

可以朴素地将用户消息简化为短文本，使用 Bag-of-Words 形式描述，即利用分词方法，将用户文本转变成词袋以描述用户特征。然而，在用户描述中，比如不具备深层含义的单词"like"使用频繁，而代表性单词"football"比较稀疏，因此不能有效地计算词袋之间的相似度，而是针对每一个特征单词，构建单词表述特征向量，其计算公式如下所示：

$$P_W(w,c_i)=\sum_{k=1}^{n}B(u_k,w)\frac{\text{like}(u_k,c_i)}{\sum_{c\in C}\text{like}(u_k,c)} \tag{9-32}$$

这里，$B(u_k,w)$表示用户是否拥有该特征 w，如果有，则为 1，否则，为 0。为了评价一个用户喜欢消费媒体中某一个类别的程度，计算方式如下所示：

$$P_{TW}(c_i|u)=\sum_{w}^{W}\text{tfidf}(w)\,P_W(w,c_i) \tag{9-33}$$

其中，W 是指用户特征实体。上述公式中将用户短文本以词袋模型表示，然而正如第 8 章中所述，词袋模型无法理解词义，而且会丢失词序等有价值的信息。因此，可以使用基于知识库理解的用户特征建模方式，将用户文本类数据使用知识库理解来构建用户概念特征向量。其中，用户概念特征以向量方式表述。同样地，将每一类概念作为一个独立单元学习概念所对应的特征，其计算公式如下所示：

$$P_I(i,c_j)=\sum_{k=1}^{n}B(u_k,i)\frac{\text{like}(u_k,c_j)}{\sum_{c\in C}\text{like}(u_k,c)} \tag{9-34}$$

其中，当用户 u 拥有兴趣 i 时，$B(u,i)$等于 1，否则，为 0。同样地，为了评价目标用户对消费媒体中某一个类别的偏好程度，可以利用如下计算方式：

$$p_{TI}(c_j|u)=\sum_{i}^{I}\text{tfidf}(i)\,P_I(i,c_j) \tag{9-35}$$

9.3.3 基于消费网络属性的用户偏好预测

不同的消费网络具有各自不同的特性。例如，社交网络提供用户一个相互交流的平台，可以发布信息，以及分享好友信息等。视频网络提供用户观看和分享视频的功能。当预测一个用户在一个消费媒体偏好时，应该考虑该媒体特征。可以使用最为朴素的预测方式推送相关信息给用户。

$$p_P(c_i|u)=\sum_{k=0}^{n}\frac{\text{like}(u_k,c_i)}{\sum_{c}^{C}\text{like}(u_k,c)} \tag{9-36}$$

9.3.4 预测模型

前面 3 小节中，分别基于用户显式反馈、隐式反馈、媒体本身特征等因素，做出用户偏好预测。本节中，对不同因素给予不同权重，得到一个综合结果，公式如下所示：

$$p(c_i|u)=\alpha p_P(c_i|u)+(1-\alpha)(\beta p_D(c_i|u)+(1-\beta)p_{T?}(c_i|u)) \tag{9-37}$$

其中，α 和 β 表示权重，并且$p_{T?}(c_i|u)$是$p_{TW}(c_i|u)$或者是$p_{TI}(c_i|u)$。

9.4　强关联预测方法实例分析

9.4.1　数据集

本实例使用用户在社交网络中的信息对用户在消费媒体的消费行为进行预测，本节首先将对研究使用的样本数据详细分析。为了研究用户社交网络的信息对在线媒体行为的预测，采集真实 Google 用户数据，用户社交网络信息来自于 Google＋，消费媒体行为来自于 YouTube。Google＋作为社交网络，被 Google 用于增加旗下众多服务之间的社交层面，在 2011 年启动。与普通社交网站相似，用户可以在 Google＋上发布消息，添加好友及与好友互动等。基于 Google 的众多用户，Google＋在 2013 年就已经成为世界上第二大社交网络，仅次于 Facebook。而 YouTube 是 2005 年注册的一个视频分享网站，支持用户上传、观看以及分享视频或短片，目前已成为最大的在线视频网站。

Google＋和 YouTube 作为 Google 旗下的产品，支持用户使用同一个账号登录使用，这就为使用用户社交网络信息对用户在在线媒体中的偏好进行预测提供了海量可信的基础数据。本实例基于这种优势，随机采集 13 956 个 Google＋用户，这些用户不仅在 Google＋上分享了 YouTube 的链接信息，给出相对完整的个人信息，而且还在 YouTube 上点击收藏过视频。每一个用户的基本信息主要包括以下两方面内容：用户 Google＋的基本个人信息，如年龄、性别、自我描述等；用户 YouTube 喜好的视频以及视频的详细信息等。其中，数据量如表 9-1 所示。

表 9-1　实验数据量

用户数	13 956
社交网络用户资料数	13 956
YouTube 视频分类数	18
用户喜好视频数	675 496

基于用户 Google＋的信息，分别对用户性别分布和感情状态分布进行统计。如图 9-7、图 9-8 所示，Google＋中男性占 60.34%，女性占 38.34%；在感情状态中，占多数的是 Single，其次是 Married，一定程度上说明用户较为年轻。这与 Google 发布的数据，即男性用户占 62%，女性用户占 37%等信息基本保持一致。

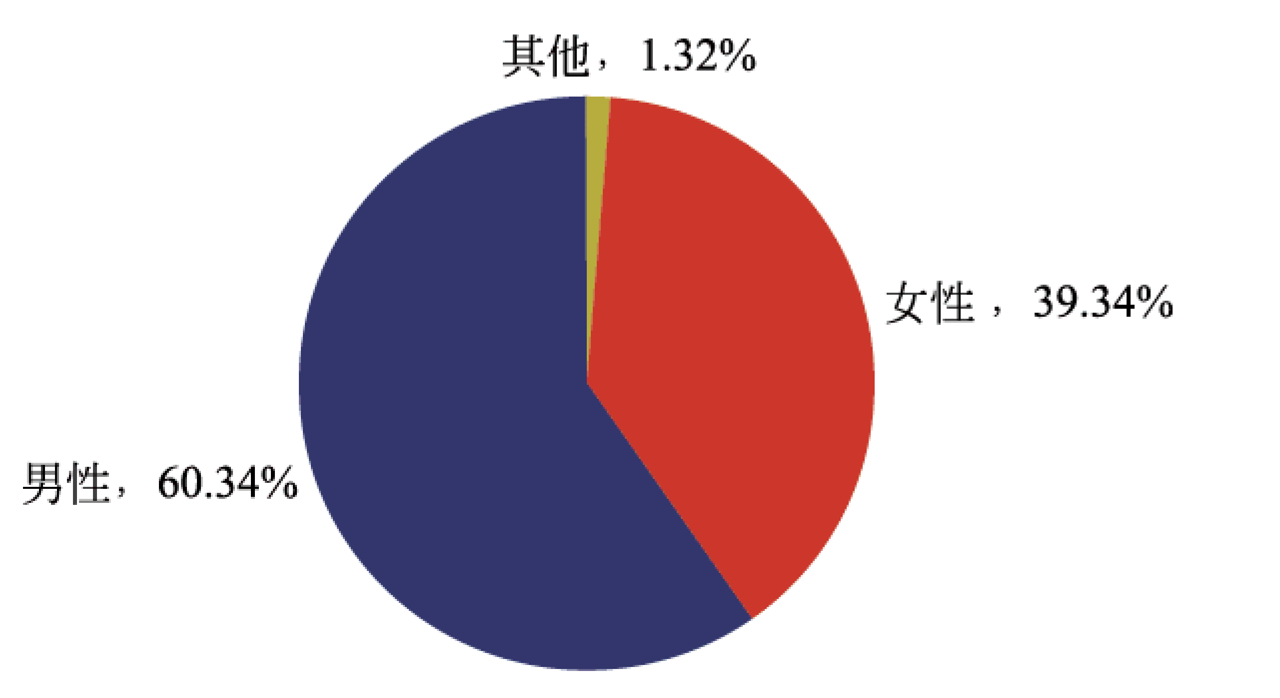


图 9-7　Google＋用户情感状态分布

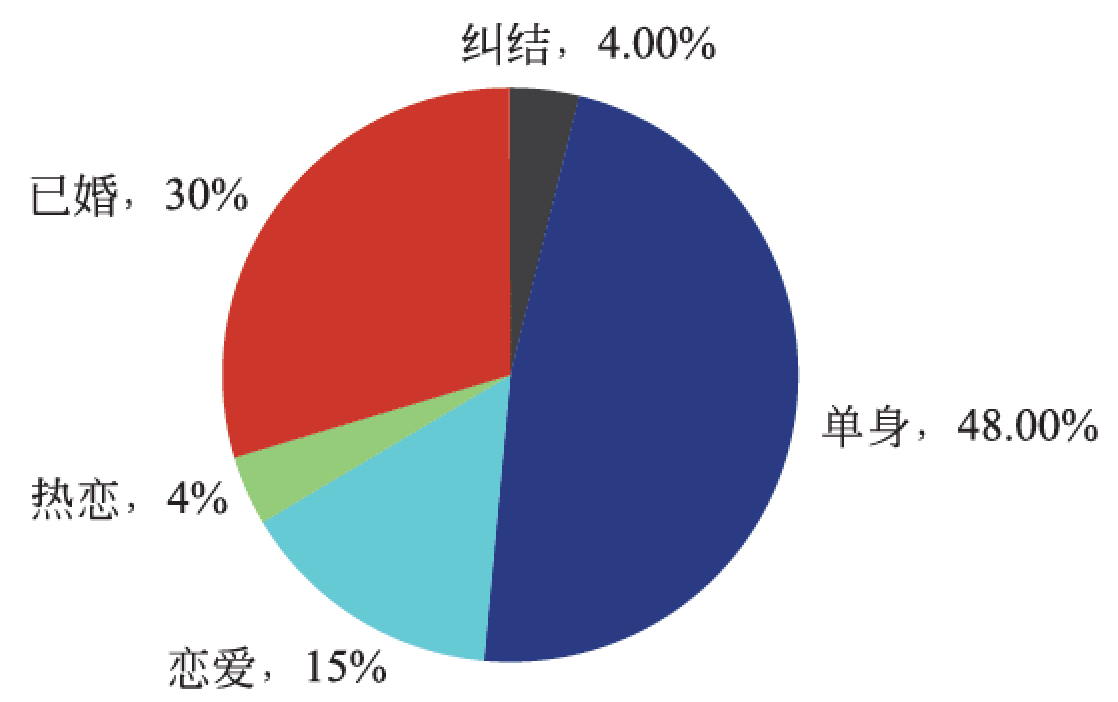


图 9-8　Google＋用户性别分布

图 9-9 统计所有用户在 YouTube 喜好视频的主要分布，可以看出，最为流行的是 Music 类，其次是 Entertainment，前两类占比 52%，所以使用流行度预测新用户的偏好可以带来比较高的准确率。对用户在 YouTube 上的行为进行分析，结果如图 9-10 所示。图 9-10 展示出用户在 YouTube 喜好视频的数量，可以看出，少量的用户分享较多的视频，而大多数的用户分享少量的视频。用户的行为符合长尾(power law)分布。

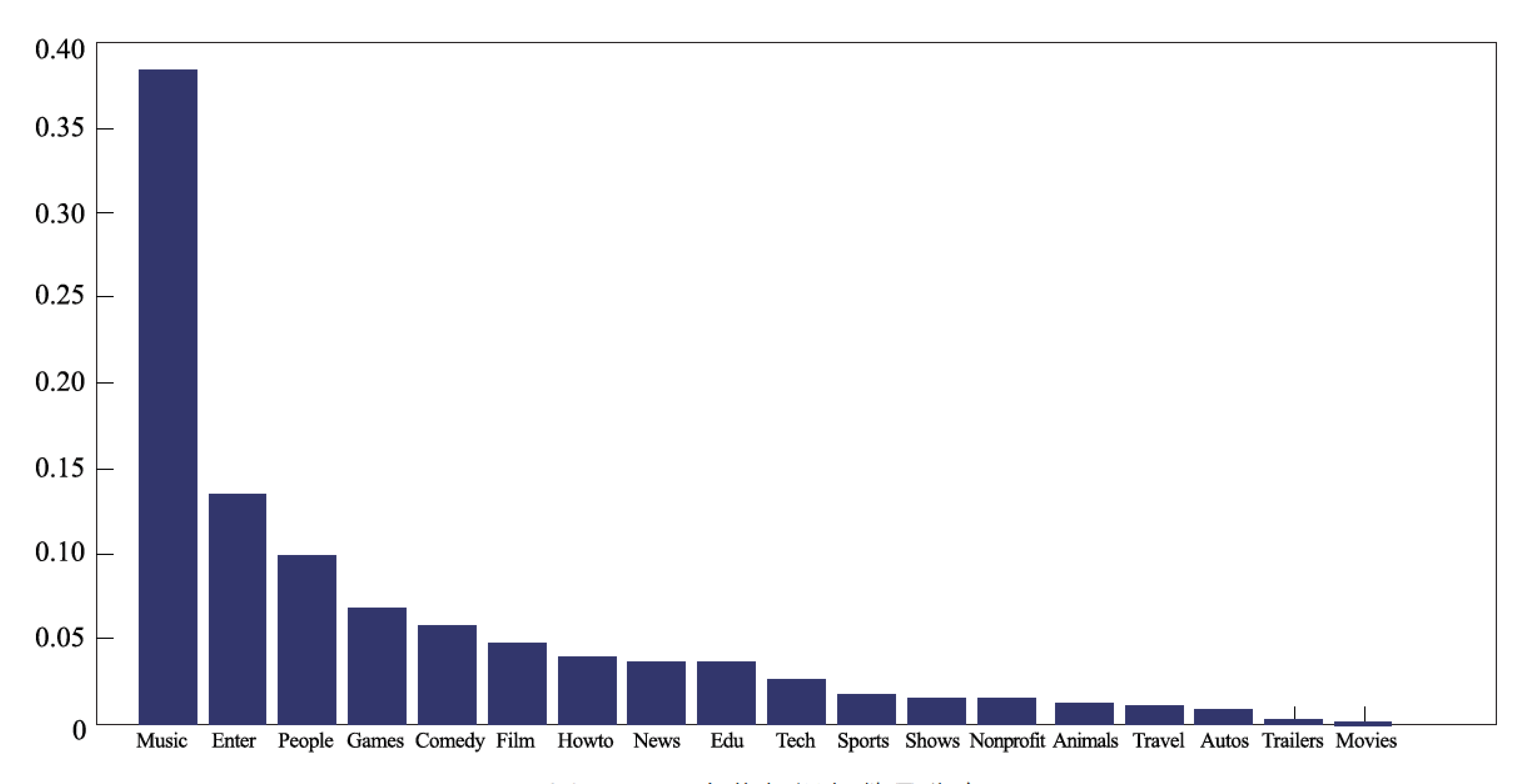


图 9-9　用户偏好视频数量分布

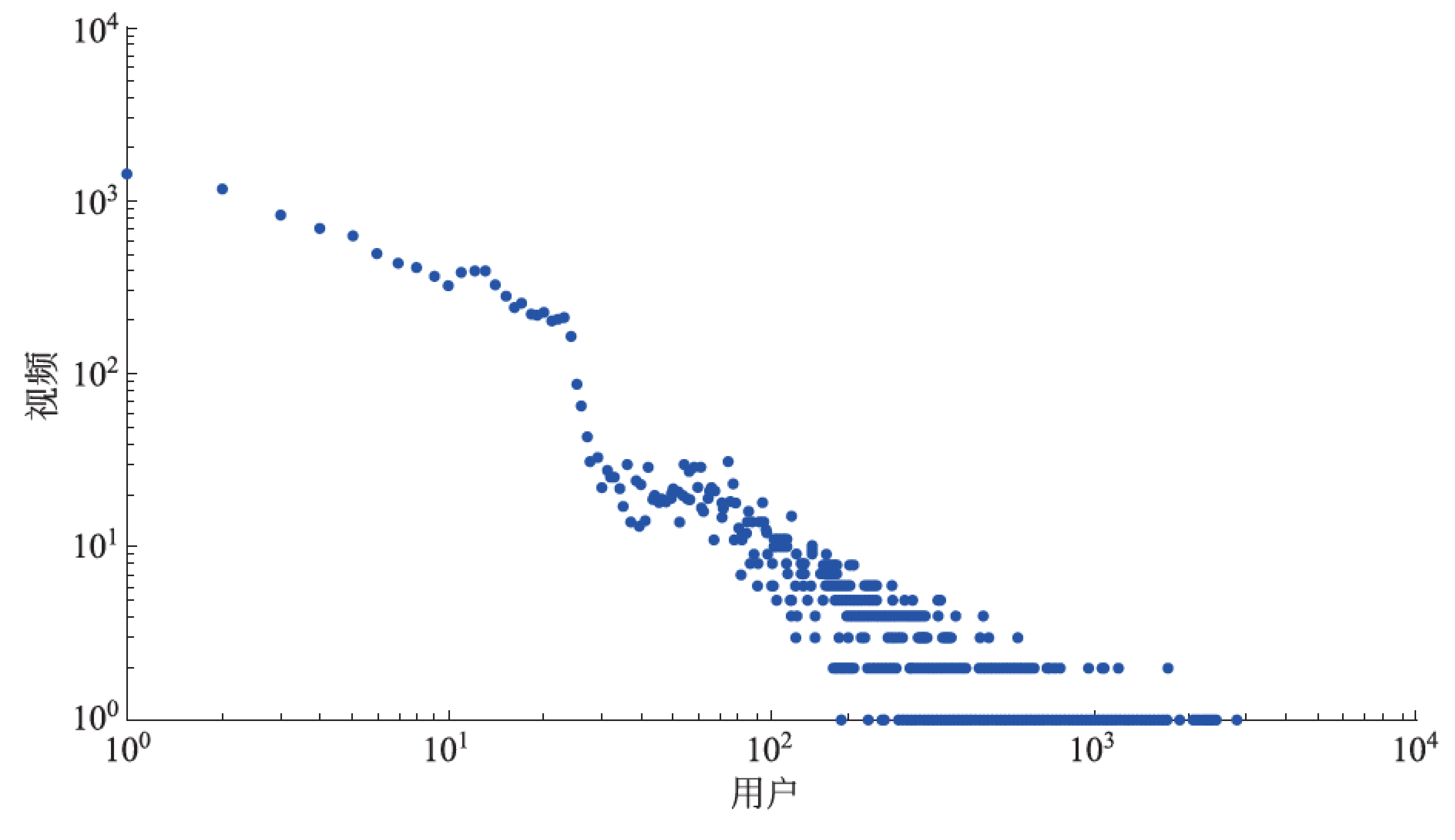


图 9-10　YouTube 视频偏好分布

9.4.2　评价指标与预测方法

本节中，将使用以下两类指标对预测结果进行评价。

准确率(P_k)、召回率(R_k)、F 指标：在预测结果的前 k 个分类中，准确率用于描述用户对预测分类感兴趣的可能性，定义为预测列表中用户真实偏好类别与所有预测类别的比率，

即 $P=\frac{tp}{tp+fp}$；召回率用于描述一个用户喜欢的产品被预测的概率，定义为预测列表中用户真实偏好的类别与用户所有偏好类别的比率，即 $R=\frac{tp}{tp+fn}$；F 指标用以综合考虑准确率和召回率，定义为 $F=\frac{2PR}{P+R}$。用户真实数据与预测数据分布矩阵可见表 9-2。

表 9-2 用户真实数据与预测数据分布矩阵

	预测类别	未预测类别
用户偏好类别	tp	fn
用户不偏好类别	fp	tn

Normalized Discounted Cumulative Gain(NDCG)：DCG 是一个衡量搜索引擎算法的指标。该指标通过对逐个返回结果打分，从而对搜索引擎或相关程序的有效性加以度量。其特点是，越相关的文档出现的结果列表位置越靠前、越好，相关性高的文档要比低的文档效果好。其中，定义 DCG 的公式如下所示：

$$\mathrm{DCG}_k=\sum_{i=1}^{k}\frac{w(i)}{\log(i+1)} \tag{9-38}$$

$w(i)$为用户偏好类的权重，计算方式如下所示：

$$w(u,i)=\frac{c_i}{\sum_{j=1}^{k}c_j} \tag{9-39}$$

其中，c_i是用户喜好 i 类视频的数量。如果权重 $w(u,i)$越大，说明用户 u 喜好 i 类视频越多，就可以使用权重 w 对用户偏好进行排序。最理想的预测结果是不仅预测出用户的偏好，而且给出的偏好排序与用户真实偏序相同。

为了计算NDCG_k，还需要计算IDCG_k，IDCG_k与DCG_k相同，不过权重是由大到小的一种理想排序。得到一个用户的DCG_k和IDCG_k后，即可算出NDCG_k，其中，

$$\mathrm{NDCG}_k=\frac{\mathrm{DCG}_k}{\mathrm{IDCG}_k} \tag{9-40}$$

我们提出通过结合异质平台数据的方式提高消费媒体预测准确性的个性化服务方法，并且结合知识库解决用户特征建模。在实验中，作为比较，提出 3 种预测策略，如下所示。

仅使用媒体特征的预测方法(S1)。

$$\alpha=1 \tag{9-41}$$

不仅使用媒体本身特征，还包括用户社交网络信息，但用户社交网络信息使用词袋模型表示(S2)。

$$p_{T?}(c_i|u)=p_{TW}(c_i|u) \tag{9-42}$$

使用媒体特征，并且包括用户社交网络特征，而这些特征是借助知识库理解的(S3)。

$$p_{T?}(c_i|u)=p_{TI}(c_i|u) \tag{9-43}$$

9.4.3 实验结果与分析

在实验中，可以通过调整参数 α 和 β 以获得最好的性能，以 0.01 的幅度进行调整，调整

空间为[0，1]。结果表明，利用媒体本身特征，即流行度预测，有很好的效果，在策略 $S2$ 中，参数 α 达到 0.5。这个结果表明，媒体本身特性是很重要的一个指标。并且从图 9-9 中可以看出，音乐类占比 38%，因为音乐类视频短小，而且内容丰富，其娱乐性较强，易被用户喜欢分享。另一方面，参数 β 占比并不大，这是可以理解的，因为无结构的用户描述提供了比用户性别、年龄更丰富的信息。更进一步地，在策略 $S3$ 中，参数 α 和 β 的值均要小于策略 $S2$ 中的值，这也表明，通过知识库对用户理解可以带来更好的效果。

使用性能最好的参数，计算结果，并对结果采用准确率和 NDCG 两个指标来评价预测结果。从图 9-11 和图 9-12 中可以看出，尽管基于媒体本身特性的预测结果可以得到一个优秀的结果，但策略 $S2$ 和 $S3$ 的结果又优于策略 $S1$，这说明，异质网络的信息对促进预测性能是非常有效的。特别是，当对用户个人描述利用知识库概念化后，可以得到一个更好的结果。

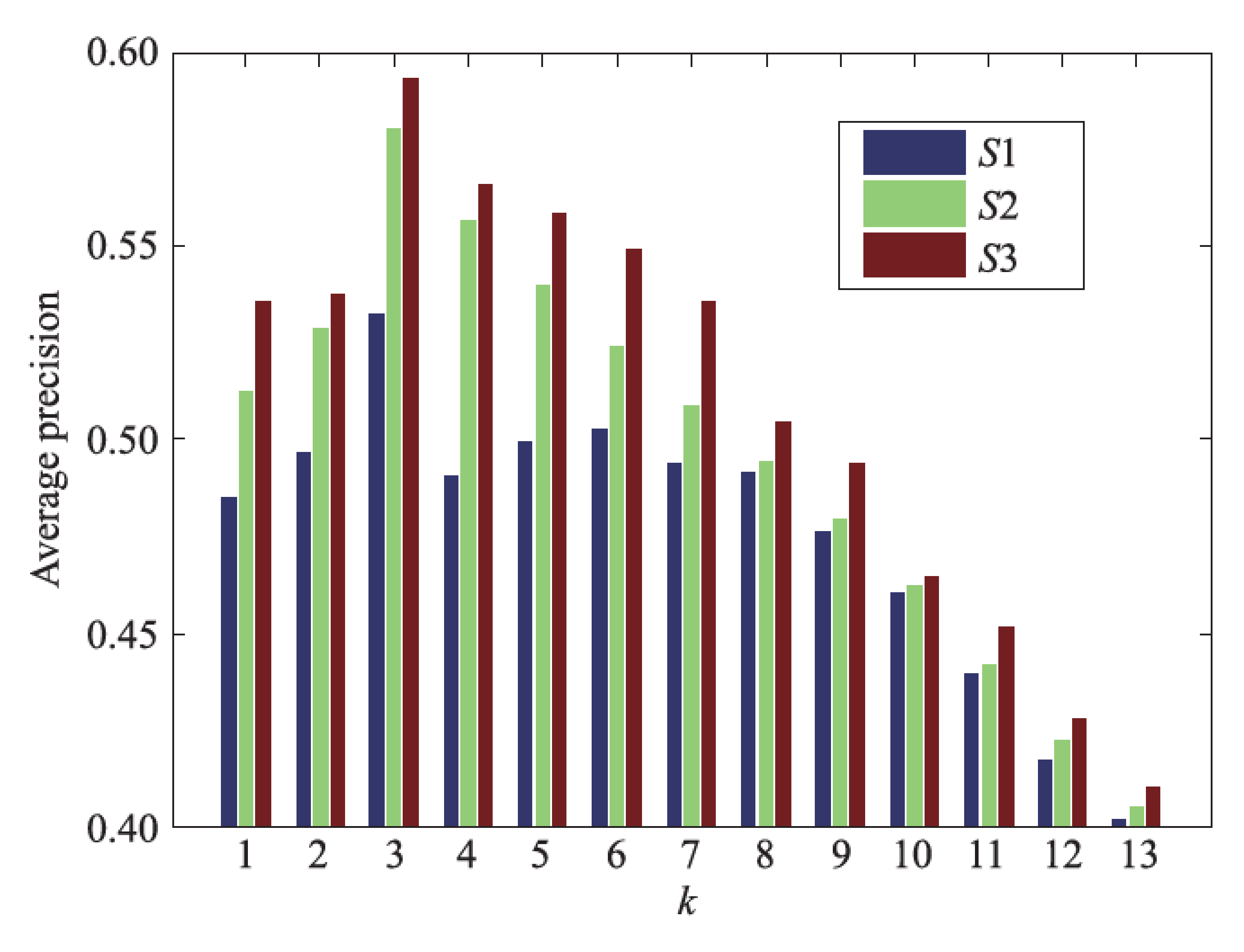


图 9-11 预测方法平均准确度

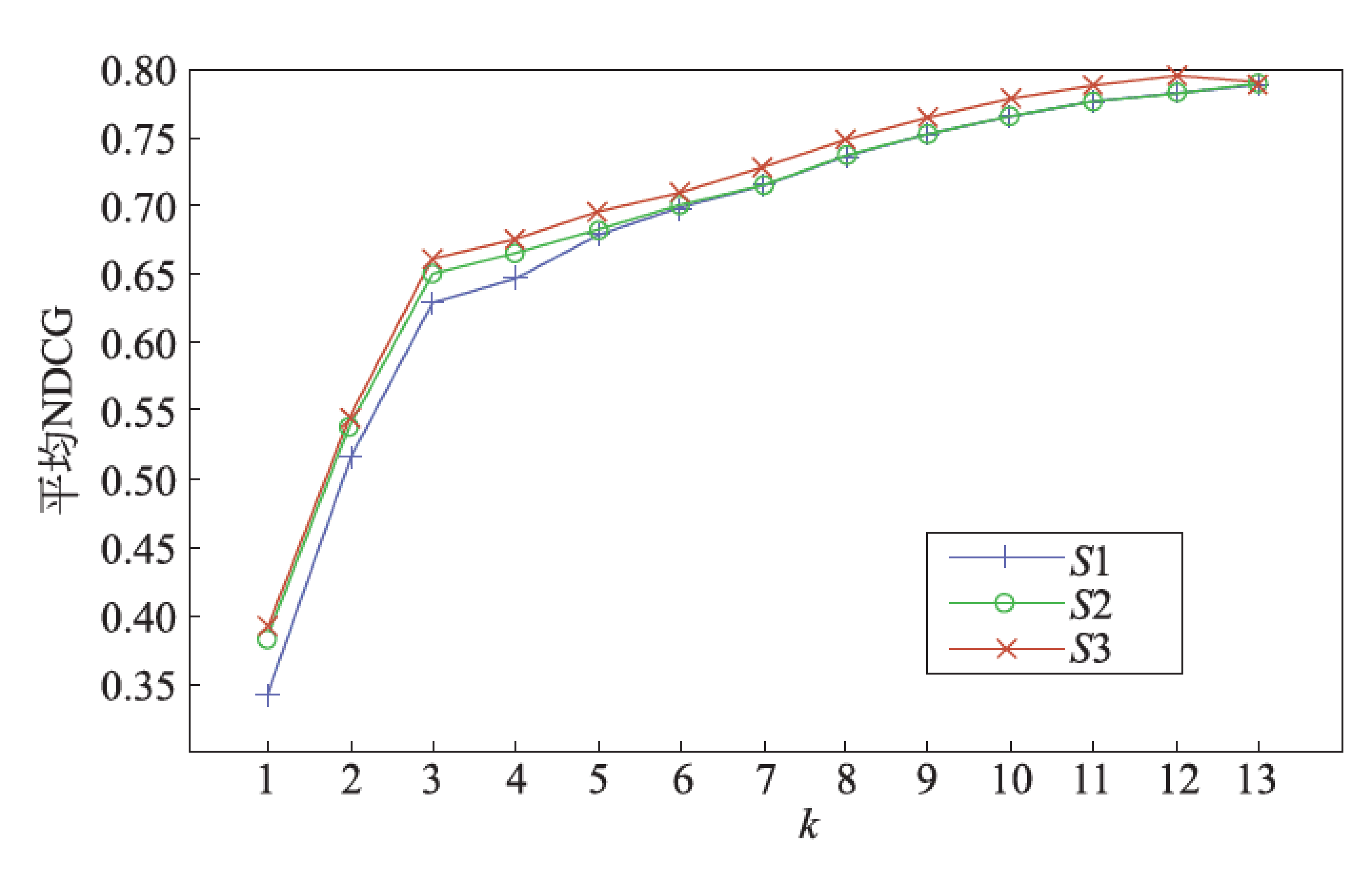


图 9-12 预测算法 NDCG 计算性能

如上文所示，超过 51%的用户喜欢的类别少于 5 个类别，因此越多结果预测，准确度越低，但是召回率越高。因此，使用 F 值来表示预测结果。由于视频类别分布的不均匀性，在 k 比较小时，预测准确性并不是特别优秀，但是随着类别的增加，用户个性化和多元化逐渐体现出来，预测准确性提高。因此，当 k 越大时，结果越好。在 k 等于 5 时，结果达到峰值。

另一方面,借助用户社交网络信息的预测使得预测结果更加稳定。实验结果还表明,使用知识库对用户偏好概念化后,可以更好地理解学习用户特征,结果比词袋模型的效果更好。

小结

本章基于所构建的用户特征,提出融合异质社交网络数据的消费媒体偏好预测方法,并在 Google+和 eBay 真实数据上给出一组实验。

首先利用构建的用户特征,基于前人的研究成果,针对不同的用户特征,给出合适的预测方法,如显式反馈特征,利用典型的基于用户的协同过滤算法;如隐式特征,利用隐语义矩阵分解模型。最后,融合上述方法,利用随机梯度下降算法,学习最优权重分配。

在此基础上,采集 Google+和 YouTube 真实数据,构建实验。利用 20 956 个 Google+和 eBay 用户数据证明,基于异质网络信息可以有效地促进预测性能,特别是将用户信息使用知识库概念化后,可以得到一个 8%左右的提升性能。

本章中,基于用户社交网络和消费媒体账号绑定的条件,利用用户在异质社交网络中的特征,预测用户在消费媒体中的偏好。然而,多数情况下,异质网络账户绑定通常不能获取,因此,在下一章中,将打破账户绑定壁垒,首先利用用户显式特征和部分隐式特征,匹配异质网络账户。在匹配基础上,利用匹配账户的社交网络特征,预测用户在消费媒体中的偏好。

参考文献

[1] Lops P, De Gemmis M, Semeraro G. Content-based recommender systems: state of the art and trends. In: Recommender Systems Handbook. Springer US, 2011. 73-105.

[2] Lin C, Kuo TT, Lin SD. A content-based matrix factorization model for recipe recommendation. In: Advances in Knowledge Discovery and Data Mining. Springer Int'l Publishing, 2014. 560-571.

[3] Wei S, *et al*. Collaborative filtering recommendation algorithm based on item clustering and global similarity. In: Proc. of the 2012 5th Int'l Conf. on Business Intelligence and Financial Engineering (BIFE). IEEE, 2012.

[4] Cai Y, *et al*. Typicality-based collaborative filtering recommendation. IEEE Trans. Knowl. Data Eng., 2014, 26(3): 766-779.

[5] Qiu L, Benbasat I. A study of demographic embodiments of product recommendation agents in electronic commerce. Int'l Journal of Human-Computer Studies, 2010, 68(10): 669-688.

[6] Liang TP, *et al*. A semantic-expansion approach to personalized knowledge recommendation. Decision Support Systems, 2008, 45(3): 401-412.

[7] Kamahara J, *et al*. A community-based recommendation system to reveal unexpected interests. In: Proc. of the 11th Int'l Multimedia Modelling Conf., 2005.

[8] Chen W, *et al*. A hybrid recommendation algorithm adapted in e-learning environments. World Wide Web, 2014, 17(2): 271-284.

[9] Klašnja-Milićević A, *et al*. E-Learning personalization based on hybrid recommendation strategy and learning style identification. Computers & Education, 2011, 56(3): 885-899.

[10] Balabanović M, Shoham Y. Fab: content-based, collaborative recommendation. Communications of the ACM, 1997, 40(3): 66-72.

[11] Rokach R, Shapira B, Kantor PB. Recommender Systems Handbook. Vol. 1. New York: Springer-Verlag, 2011.

[12] Latch EK, *et al*. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conservation Genetics, 2006, 7(2): 295-302.

[13] Dumais ST. Latent semantic analysis. Annual Review of Information Science and Technology, 2004, 38(1): 188-230.

[14] Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. Journal of Machine Learning Research, 2003, 3: 993-1022.

[15] Berger AL., Della Pietra VJ, Della Pietra SA. A maximum entropy approach to natural language processing. Computational Linguistics, 1996, 22(1): 39-71.

[16] Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering. In: Proc. of the 24th Int'l Conf. on Machine Learning. ACM, 2007.

[17] Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters, 1999, 9(3): 293-300.

[18] Paterek A. Improving regularized singular value decomposition for collaborative filtering. In: Proc. of the KDD Cup and Workshop. 2007.

[19] Liu Q, *et al*. Predicting user likes in online media based on conceptualized social network profiles. In: Web Technologies and Applications. Springer Int'l Publishing, 2014. 82-92.

[20] Ruiz EV. An algorithm for finding nearest neighbours in (approximately) constant average time. Pattern Recognition Letters, 1986, 4(3): 145-157.

[21] Friedl MA, Brodley CE. Decision tree classification of land cover from remotely sensed data. Remote Sensing of Environment, 1997, 61(3): 399-409.

[22] McCallum A, Nigam K. A comparison of event models for naive bayes text classification. In: Proc. of the AAAI-98 Workshop on Learning for Text Categorization, Vol. 752. 1998.

[23] Wang SC. Artificial neural network. In: Interdisciplinary Computing in Java Programming. Springer US, 2003. 81-100.

[24] Wagstaff K, *et al*. Constrained *k*-means clustering with background knowledge. In: Proc. of the ICML, Vol. 1. 2001.

[25] Zhang S, *et al*. Learning from incomplete ratings using non-negative matrix factorization. In: Proc. of the SDM, Vol. 6. 2006.

[26] Mnih A, Salakhutdinov R. Probabilistic matrix factorization. In: Advances in Neural Information Processing Systems. 2007.

[27] Singh AP, Gordon GJ. Relational learning via collective matrix factorization. In: Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. ACM, 2008.

第 10 章
社交网络用户在弱关联消费网络中的偏好预测

第 9 章利用异质社交网络用户数据预测用户在消费媒体中偏好的方法，结果表明，该方法可以有效地预测社交网络用户在消费网络中的行为。但需要以用户的社交网络账户与消费媒体账户绑定为条件，即确定两个账号同属于同一个个体。然而，现实中，由于媒体信息垄断以及用户信息隐私性，账户绑定并不能显而易见地获得。在本章中，利用用户在社交网络和消费媒体中的特征匹配用户账户，并将学习匹配的账户关联称作弱关联。在弱关联的基础上，利用社交网络特征预测用户在消费媒体中的偏好。

用户账号匹配是解决很多有意义问题的前提条件，已被很多学者研究。本章基于前人的研究，扩展博弈论中 stable matching 的思想，对异质网络用户账号进行匹配，并给出了一组实验证明。

10.1 概述

稳定匹配作为一种博弈算法，已被计算机领域专家和经济学家深入学习。稳定匹配问题源于男性和女性稳定婚姻问题，如图 10-1，图 10-2 所示。形式化定义为，有男士集合 M 和女士集合 W，对于集合 M 中每一位男士m_a及集合 W 中的每一位女士w_b，都有一个严格不等的偏好

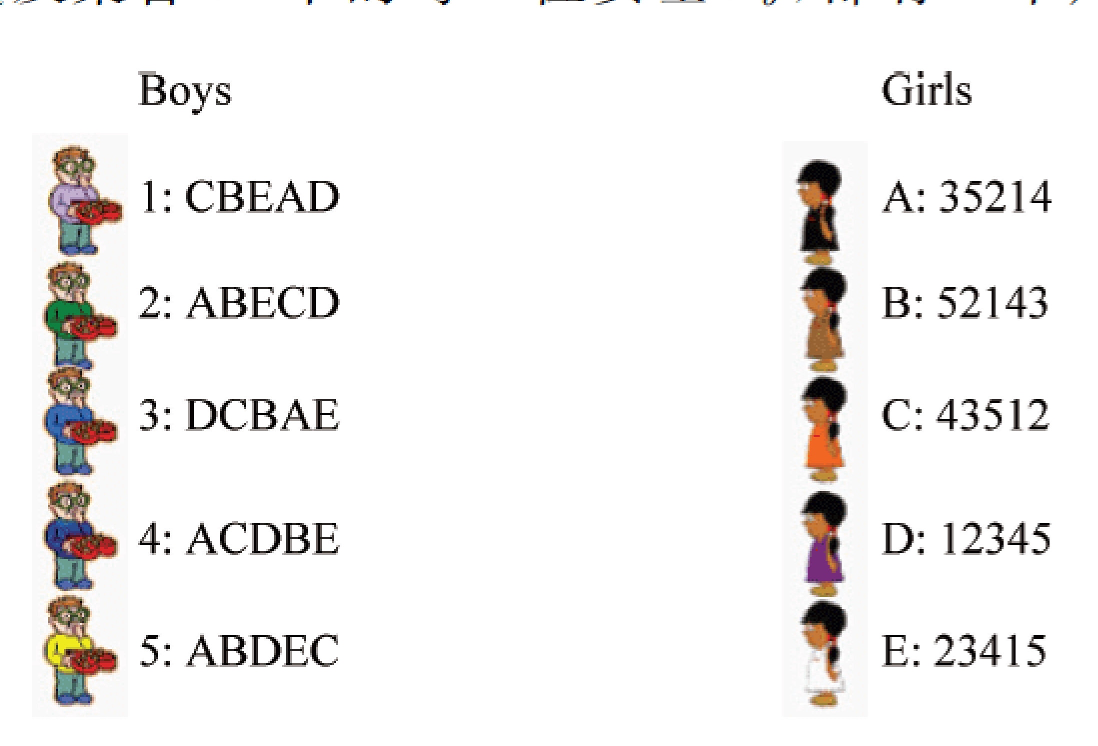


图 10-1 初始用户偏好分布

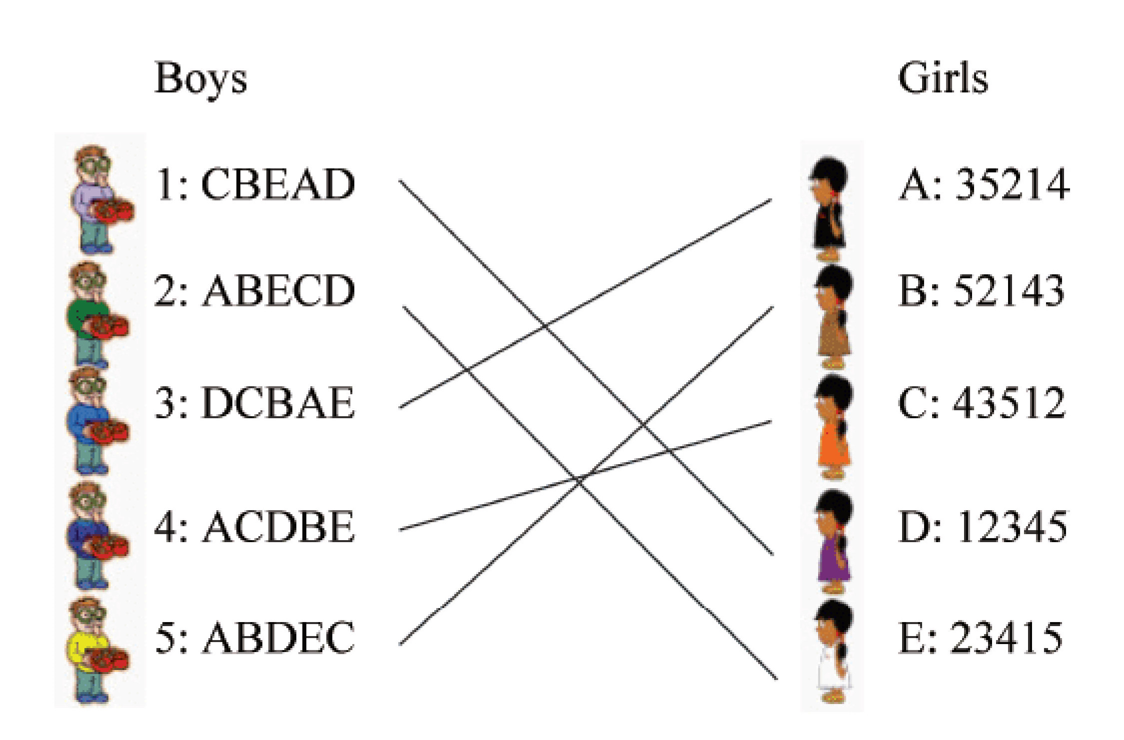


图 10-2　基于稳定匹配的计算结果

程度。同样，对于集合 W 中每一位女士 w_c 及对每一个集合 M 中的男士 m_d，都有一个严格不等的偏好程度。对于个体的偏好程度，可以使用 $a \succ_u b$ 来表示，表示 u 偏好 a 要比偏好 b 程度高。一个稳定匹配就是对每一位男士 m 和每一位女士 w，都能找到偏好程度最高的对方。一个匹配是稳定的，当然匹配结果是稳定的。大量文献表明，稳定匹配是肯定存在的，但是由于切入角度不同，存在男士稳定匹配和女士稳定匹配的结果。

原始稳定匹配问题中，对每一个节点，要求要有严格不等偏好程度。并且，对于双方，其偏好的程度可以不同。然而，在账号匹配过程中，比如社交网络账号和消费媒体账号，这两个条件并不能满足，偏好程度由两个账号相似程度计算，可能不是严格不等的，并且两个账号相似度是完全相同的。

假如存在一个社交网络 G^S 和一个消费媒体 G^E，并且假设，用户在社交网络 G^S 中拥有丰富的个人信息和充分的历史行为信息，而在消费媒体 G^E 中，用户个人特征可能不完整，但是拥有大量的历史行为信息。$G^S=(U,F^S,M^S)$ 表示用户在社交网络中的特征，其中，$U=\{u_1, u_2,\cdots,u_n\}$ 表示用户账号集合，$F^S=\{f_1,f_2,\cdots,f_k\}$ 表示用户显式特征，即人文特征信息，比如用户昵称、用户年龄等。而 M^S 表示用户历史行为信息，在社交网络中，用户历史行为通常是文本，可以使用词袋模型表示，也可以通过第 3 章使用知识库将用户特征概念化，得到用户兴趣向量。同样地，用户在消费媒体中可以表示为 $G^E=(U,F^E,H^E)$，其中，$U=\{u_1, u_2,\cdots,u_n\}$ 表示相同的用户集合。但在消费媒体中，用户人文特征可能很少，而历史行为信息较为丰富。

用户账号匹配的定义就是，将社交网络 G^S 账号和消费媒体 G^E 账号中属于同一个个体的账号关联起来。

10.2　相关研究工作

前面对个性化预测的模型与方法展开了详细的讨论，其中融合多资源的模型可以融合多异质平台数据，提高预测的准确度。但在这种技术中，需要异质网络账户是绑定的、显式的。然而现实情况下，由于媒体的局限性以及用户的隐私性，通常很难获取账户之间的绑定，而异质网络账户匹配可以有效地解决账户关联问题。现实中，异质网络账户匹配是解决很多重要问题的前提和必要条件，比如利用异质网络信息进行用户好友推荐。因此，很多学

者对异质网络账户匹配做过深入学习，本节将针对异质网络账号展开学习综述。异质网络账号匹配方法通常可以分为 3 类：无监督的匹配(unsupervised matching)[1-3]、有监督的匹配(supervised matching)[4-7]以及基于博弈思想的匹配(game-theoretic matching)[8]。

无监督匹配假设用户在异质网络中拥有相似的显式反馈特征。比如使用相似的用户名、相同的性别和地理位置、甚至拥有相似的好友关系。因此，基于无监督的匹配可以划分成两类：基于语义特征(semantic based)和基于结构特征(graph based)。基于语义的无监督匹配方法判断两者人文特征的相似程度，越相似，说明越可能是同一个用户。Buccafurri[1]证明用户喜好使用相同或者相似的昵称。Carmagnola[2]延伸了使用用户个人资料的方法，并对不同的内容赋予不同的权重，基于最终的加权和，验证是否属于同一个个体。然而，不同的在线媒体，账户资料内容差异很大，比如社交网络需要用户填写年龄、性别、偏好标签等内容，而购物媒体，比如 eBay，并不需要用户填写这些信息，这就造成基于语义无监督的学习方法无法应用。基于结构的无监督匹配通常用于异质社交网络中，如假设用户好友关系保持不变，则可以利用用户好友关系结构图验证两个账户的相似程度。Bartunov[3]证明用户好友关系在账户匹配中所起的重要作用，并证明这个属性要优于基于账户语义属性的匹配方法。无监督匹配方法充分利用用户显式特征，其计算方法相对直接，易于理解。然而，在多数网络中，用户显式特征信息较少，甚至不存在，比如用户好友仅仅显式存在于社交网络中，因此多数情况下无监督方法不适用；其次，用户填写个人信息通常比较随性，每次注册使用信息差异可能比较大，无监督匹配方法误差比较大；最后，当用户数量增多时，用户显式特征碰撞率变大，造成无监督匹配方法的准确率降低。

有监督的学习方法，除了利用用户显式特征外，更多地考虑的是用户隐式反馈。用户同一时间在异质网络的表现行为，通常存在相似性或者关联性，因此可以利用这些隐式反馈匹配异质账户。基于监督的匹配主要源于 link prediction 的研究，试图从用户历史行为中学习用户特征，并对特征相似的用户进行映射。Backstrom[4]提出一种随机游走的监督学习算法，估算边的权重。Tang[5]提出 TranFG 模型，该模型融合常用的社交网络理论和监督学习模型，可以有效地促进账户匹配准确度。相比于无监督的学习，监督学习往往能够得到更高的准确率。Peled[6]提出了一种有监督学习的账户匹配方法，并证明具有良好的表现性能，虽然不完整的信息会使学习方法准确性降低。Xu[7]证明，结构用户社会特性和个人特性可以有效地促进嫌疑人匹配。一方面，相比于用户显式特征，网络中用户的隐式特征数据量更大，信息更丰富，而随机性较小；另一方面，随着用户量的增多，显式特征冲撞率增大，但隐式反馈的特征更具有代表性。然而必须指出的是，基于监督学习的匹配需要预先学习匹配规则，计算复杂度比较高，而且会存在长尾效应，即若某一个账户信息涵盖广，账户之间相似度比较高，可能会导致多数账户映射到该账户。

近几年来，随着博弈论的深入研究，一些学者受博弈算法中收益与损失概念的启发，给出一些基于博弈算法的异质网络账户匹配方法。Kong[8]依据博弈论算法中的 Stable Matching 算法，提出异质网络的账户匹配算法。不同于监督和无监督学习算法，博弈论算法可有效均衡账户，不会存在一个账户对应多个账户的情况，解决了匹配中的长尾效应。同样地，博弈论匹配算法计算复杂性高。

10.3 社交网络与消费网络间的用户账户关联方法

10.3.1 用户账户相似度计算策略

用户账户的相似度计算可以采用多种成熟策略，下面列举了单词距离、等级差距离、兴趣向量距离这 3 种常用方法。

单词距离：很多研究统计发现，个体偏向使用相同或者相似的昵称，因此可以使用 Levenshtein 距离计算用户一些属性的相似程度，比如用户昵称。Levenshtein 距离，又称编辑距离，指的是两个字符串之间由一个转换成另一个所需要的最少编辑操作次数。编辑包括字符替换、插入字符和删除字符。Levenshtein 可以计算两个字符之间的相似程度，借此计算两个账号之间相似程度，公式如下所示：

$$W_D = 1 - \frac{d_{\text{lev}}(u_1, u_2)}{\max(\text{len}(u_1), \text{len}(u_1))} \tag{1}$$

其中，$d_{\text{lev}}(u_i, u_j)$表示 Levenshtein 距离。

等级差距离：有一些用户特征，比如用户性别、邮箱、地理位置等信息，是非常关键的特征信息。因为这些信息可以有效辨别信息的主体是否属于同一个个体。为了比较这些特征信息，采用等级差的方式计算两个账号相似度，其计算方式如表 10-1 和表 10-2 所示。

表 10-1　基于用户性别的相似度计算

Gender	Male	Female	Others
Male	1	−1	0
Female	−1	1	0
Others	0	0	1

表 10-2　基于用户年龄区间的相似度计算

Age_range	10～20	20～30	30～40	40～50	50～60
10～20	1	0.5	0	−0.5	−1
20～30	0.5	1	0.5	0	−0.5
30～40	0	0.5	1	0.5	0
40～50	−0.5	0	0.5	1	0.5
50～60	−1	−0.5	0	0.5	1

兴趣向量距离：用户历史行为是一个非常丰富的用户特征信息，可以将用户历史行为使用知识库转变成兴趣向量。但是由于媒体特征，用户兴趣向量可能完全不同，直接使用距离计算两者相似程度是不切实际的。因此可以借助训练集，对用户特征进行关系计算，然后利用训练关系计算用户账号之间的相似度。

10.3.2 用户账户匹配方法

前一节介绍了如何利用 3 种策略计算用户之间的相似度，本节基于博弈论稳定匹配算

法并结合异质网络账户匹配的特点，遵循一对一的匹配原则，提出用户账户匹配算法，如表 10-3 所示。

表 10-3　异质网络账号匹配算法

Algorithm 10.1：Account stable matching.

```
Input：two disjoint agents N^A and N^B
        Similarity between agents S;
Output：stable matching agent pairs.
begin
  pairs=∅
  repeat
    exists free n_i^A in N^A and n_j^B in N^A
      if n_i^A's most preference is n_j^B and n_j^B's most preference is n_i^A：
        Remove n_k^B's preference to n_i^A in B
        Remove n_k^A's preference to n_j^B in A
    put {n_i^A, n_j^B} into pairs
end
```

在每一次迭代中，从集合 A 中随机选择一个节点n_i^A，查询该节点对 B 集合中每个节点的偏好程度，选择偏好最高的一个，假设选中的节点为n_j^B。如果n_j^B其偏好最高的节点恰好是节点n_i^A，那么认为这一对节点就是稳定匹配的；否则，不是稳定节点，将继续选择，直到找到稳定节点。

该算法具有以下 3 个特性。

特性 1：可计算性。在每一次迭代过程中，如果存在一对节点被计算出来，那么算法可以在有限步执行完毕。

证明：假如一次计算中，没有节点被计算出来。即表示没有一组稳定匹配的节点，假设节点n_i^A的最相似点为n_j^B，然而n_j^B最相似的节点却是$n_{i'}^A$，因此，$S(n_j^B,n_{i'}^A)>S(n_j^B,n_i^A)$。基于上述算法可以看出，$S(n_j^B,n_{i'}^A)=S(n_{i'}^A,n_j^B)$以及 $S(n_j^B,n_i^A)=S(n_i^A,n_j^B)$，因此，$S(n_{i'}^A,n_j^B)>S(n_i^A,n_j^B)$，这与节点$n_i^A$是$n_j^B$最相似的点相矛盾。因此，算法可以在有限步完成。 □

特性 2：稳定性。不论从哪一个方向开始计算，其计算结果是一致的。

证明：假设结果是不稳定的，存在一对不稳定节点$\{n_i^A,n_j^B\}$，而n_i^A更偏向$n_{j'}^B$，n_j^B更偏向$n_{i'}^A$。由于$\{n_i^A,n_j^B\}$是不稳定节点，所以，$n_{j'}^B>_{n_i^A}n_j^B$。而在上述算法中，节点n_i^A将首先映射到节点$n_{j'}^B$而不是节点n_j^B，但节点n_i^A没有匹配节点$n_{j'}^B$，因为节点$n_{j'}^B$与节点$n_{i'}^A$相似度更大。因为节点n_j^B与节点$n_{i'}^A$匹配，即$n_{i'}^A>_{n_j^B}n_i^A$，该结论与$\{n_i^A,n_j^B\}$不稳定匹配相矛盾，所以不论算法首先计算哪一个节点，其计算结果都是稳定的。 □

特性 3：唯一性。不论算法切入点为何，总能给出一个唯一匹配的结果。

证明：在每一次迭代计算中，无论从任何一段切入，算法都是选择最高的相似结果。假设存在一对节点$\{n_i^A,n_j^B\}$可以从 A 端选择，而不在 B 端选择。意味着n_j^B可以选择的节点不是n_i^A，而 $S(n_j^B,n_i^A)=S(n_i^A,n_j^B)$，这与节点$n_j^B$是节点$n_i^A$最相似的节点相矛盾。因此，不论算法从哪一端切入，其计算结果都是唯一的。 □

10.3.3 用户账户关联方法实例分析

本实例采用真实的 Facebook 和 eBay 用户数据。在 Facebook 中,用户通常填写很多个人信息,所以用户 profile 是相对丰富的。除了用户 profile 信息外,Facebook 用户还不断更新和分享个人消息。而在 eBay 中,用户购买物品后通常会有卖家的反馈,通过收集卖家的反馈记录,可以收集该用户的购买历史记录。

数据集共采集到 507 个 Facebook 用户个人信息,确保这些用户在 eBay 上拥有账号,并有足够的历史信息。这些用户中,58.06%为男性用户,41.94%为女性用户,这个数据与 eBay 男女比例吻合,说明实验使用数据是具有代表性的。在用户 Facebook 数据中,共采集到 239 772 条消息,分布在 2014 年 1 月～2014 年 6 月之间。

同样,针对用户在 eBay 中的购物历史,共采集到 31 658 次购买行为,如图 10-3 所示,用户购买行为符合长尾效应,也符合实际数据。依据 eBay 的分类方式,将产品归为 36 类。

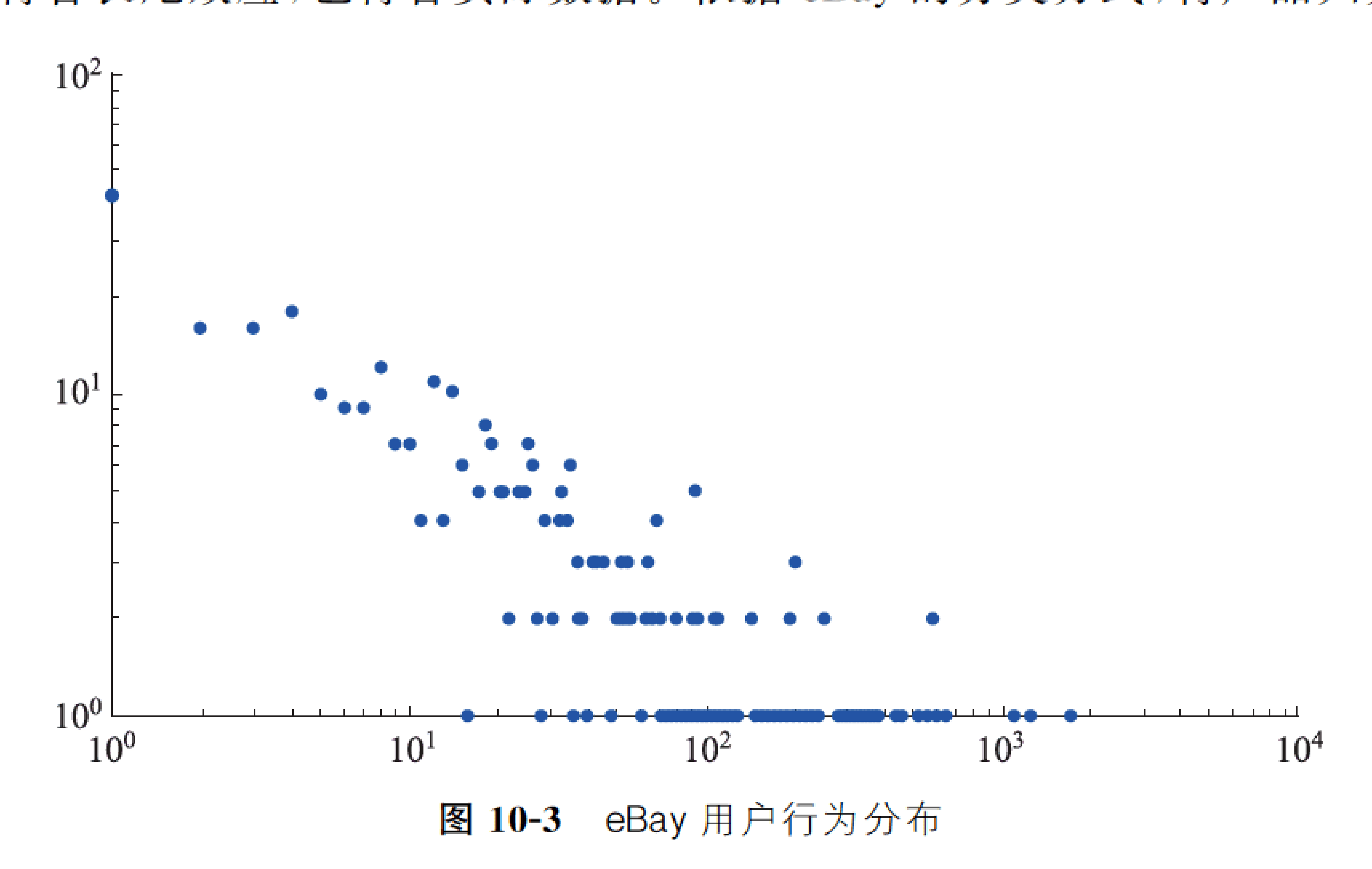


图 10-3 eBay 用户行为分布

为了验证匹配算法的性能,采用准确率评价算法性能。其中,准确率公式如下所示:

$$Acc = \frac{\mathrm{len}(U_{\mathrm{Correct}})}{\mathrm{len}(U_{\mathrm{Total}})} \tag{2}$$

其中,U_{Correct}表示匹配正确的用户数,U_{Total}表示总的用户数。

分别选用下述方法用于账号匹配,将提出的方法与其他方法进行比较,比较方法包括:

- 无监督的匹配方法(UAM)　正如前面所述,大量研究人员使用用户特征信息进行匹配,因此利用用户姓名、性别、年龄、地理位置等信息展开无监督的匹配算法实验。
- 有监督的匹配算法(SAM)　用户历史记录或者兴趣向量可有效代表一个用户特征。但是用户兴趣向量无法直接用于计算用户相似度,而是需要利用所匹配的用户进行关联特征学习,并形成用户历史行为兴趣向量。有两种表示方法:基于词袋表示(corpus based)和基于知识库概念化表示(knowledge based),所以在本次关联实验中有两种计算方法。
- 基于稳定匹配的账号匹配(HAM)　在充分利用用户的显式特征和隐式特征的基础上,借鉴并扩展稳定匹配算法用来匹配账户。同样地,用户历史特征可以包括基于词袋表示(corpus based)和基于知识库概念化(knowledge based)表示的方法。

实验中分别选用不同数量的数据测试，其中每一组数据平均分成两部分，一部分用作训练数据，另一部分用作测试数据。数据主要包括 3 种类型数据：用户昵称、用户性别和地理位置、用户兴趣向量。对每一种类型数据，使用不同的策略计算账号相似度，并给予最合适的权重因子 α，β 和 γ。权重因子区间是[0,1]，加和为 1，以 0.01 的幅度不断调节、测试以获取最优结果。

第 1 组实验结果如表 10-4 所示，可以看出，无监督算法可以达到一定准确率，这主要取决于用户是否使用相似的 profile 信息。该方法的准确率会随数据量变大而有所降低，原因是大量的数据会造成 profile attribute 相似的用户越多，从而降低用户特征的区分度。有监督的学习通过加入用户历史行为来匹配用户账号，可以使匹配成功率显著提高，说明用户历史行为可以有效地促进账号匹配。并且，当训练数据增多时效果更加显著。而 HAM 计算结果明显优于前两种计算结果，其计算结果支持本文提出的观点，即在账号匹配过程中，使用用户的历史行为并保持 one-to-one 的限制可以有效地匹配用户账号。

表 10-4　不同异质网络账号匹配算法性能

	size＝50	size＝100	size＝200	size＝300	size＝500
UAM	0.46	0.37	0.38	0.37	0.35
SAM(corpus-based interest vector)	0.50	0.38	0.41	0.38	0.36
SAM(knowledge-based interest vector)	0.55	0.38	0.42	0.39	0.38
HAM(corpus-based interest vector)	0.70	0.43	0.44	0.44	0.41
HAM(knowledge-based interest vector)	0.65	0.45	0.46	0.46	0.42

在第 2 组实验中选用用户 profile attribute (PA)，使用朴素词袋表示用户兴趣，以及使用第 8 章中基于知识库理解的用户特征来描述用户偏好，实验结果如表 10-5 所示。可以看出，profile 可以有效地进行账号匹配，但是计算结果可能会因数据的增多而有所降低，这主要是因为数据越多，相似账户越多，用户特征冲撞率变大，用户特征区分变难。使用用户兴趣向量映射用户准确率较低，主要是训练数据不够大。但是，CO 和 KN 随数据的增加效果越好，这主要是因为数据越多，模型训练的结果就越好，用户区分的效果也因此越好。同时还能看到，使用知识库概念化用户历史行为的效果明显优于 corpus 方法，因为使用知识库可以有效地对用户行为进行定义，可以更好地描述用户偏好。

表 10-5　基于不同数据的异质网络账号匹配算法性能

	size＝50	size＝100	size＝200	size＝300	size＝500
Profile attributes	0.55	0.40	0.42	0.41	0.36
Corpus-Based interest vector	0.05	0.0	0.02	0.03	0.04
Knowledge-Based interest vector	0.0	0.01	0.02	0.04	0.05
Profile attributes＋Corpus-Based interest vector	0.70	0.43	0.44	0.44	0.41
Profile attributes＋Knowledge-Based interest vector	0.65	0.45	0.46	0.46	0.42

10.4 弱关联预测实例分析

第 9 章提出了融合强关联用户社交网络特征的用户偏好预测方法，而在上一节中使用博弈论方法对异质网络账号进行匹配。本节基于前面的研究基础，基于匹配的账户以及利用融合用户社交网络特征的预测方法，给出融合弱关联账户的消费行为预测用例。本节的数据仍然使用 10.3 节中的 Facebook 和 eBay 数据。

在本实验中，FHSP 方法与仅仅采用用户历史行为的 HIST 方法以及利用用户社交网络信息的矩阵分解 MF 方法进行对比。后两种方法都是当前最为流行的方法。除此之外，为了分析冷启动问题，还采用了 POPU 作为对比方法。

从图 10-4 和图 10-5 可以看出，基于流行的预测方法，在一定程度上可以得到一个比较优秀的结果，由于 eBay 媒体的特性，用户购买必需品占据多数，比如 Clothings 占总消费的 13.08%。这也是将媒体特征加入考量范围的原因。

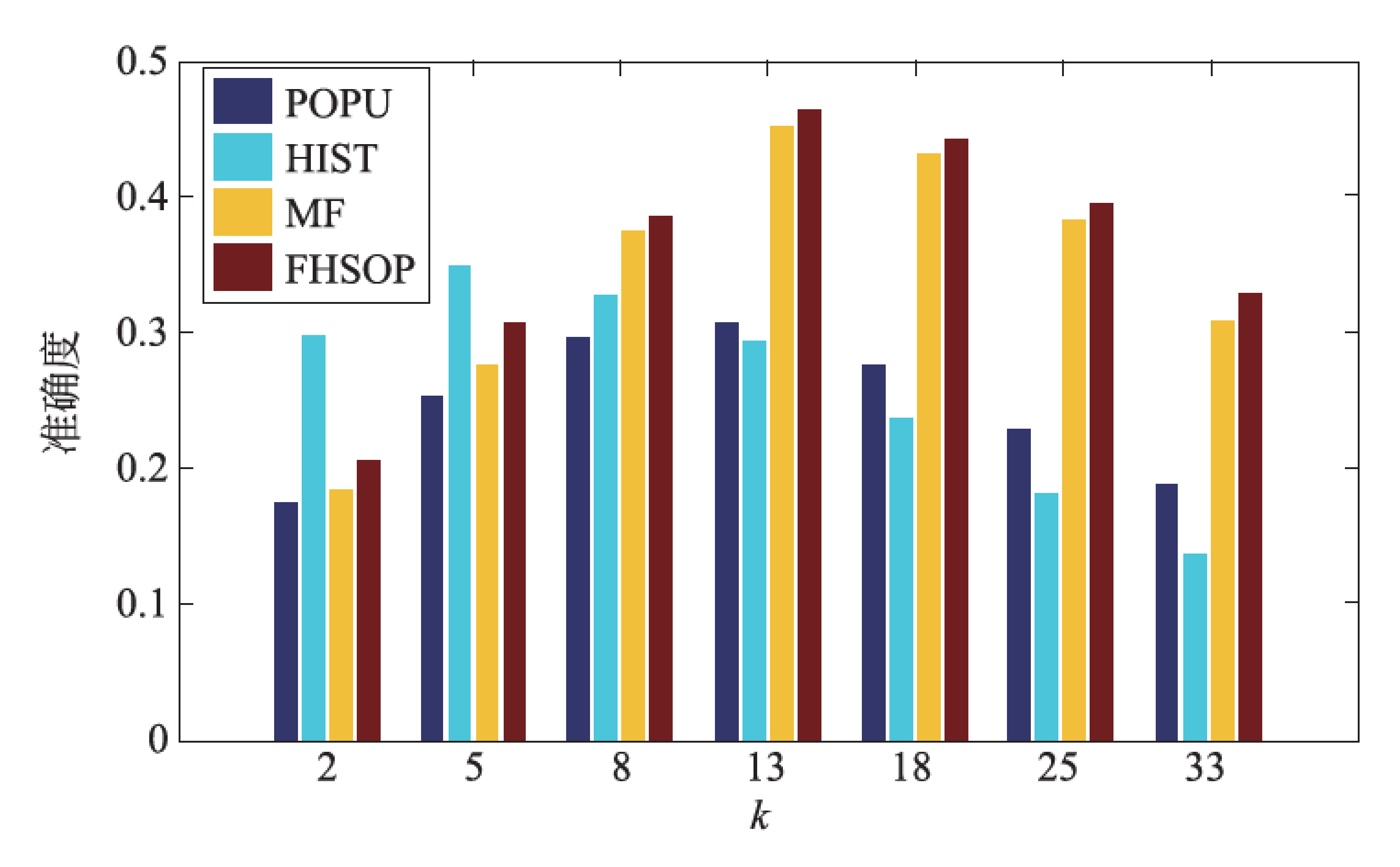


图 10-4 不同算法的计算准确度

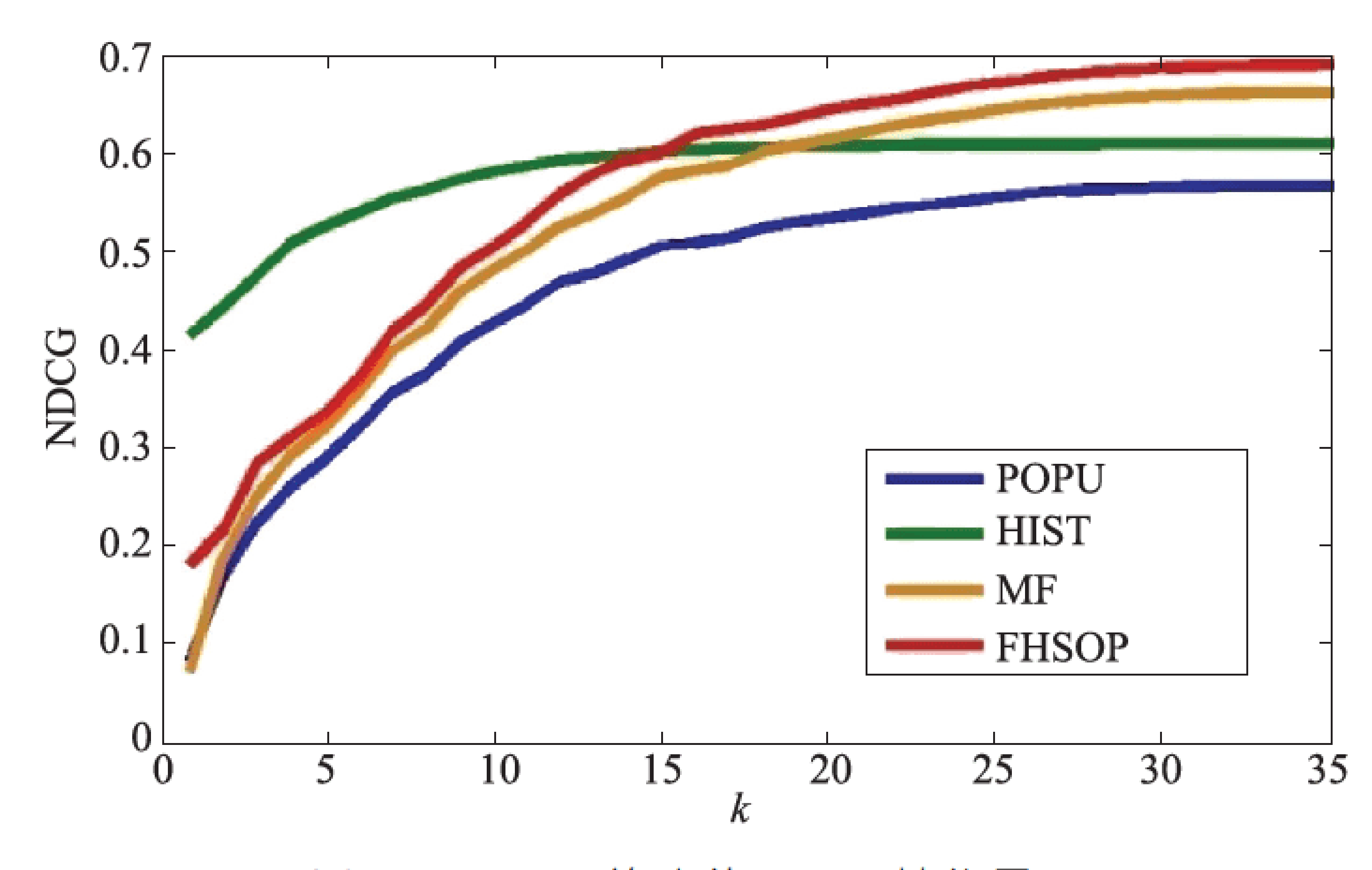


图 10-5 不同算法的 NDCG 性能展示

基于用户历史行为的预测，在 k 比较小的情况下，其效果非常好，这可以归因为用户喜欢购买已经买过的物品；但当 k 逐渐变大时，其效果变差。其原因在于，基于用户历史行为的推荐不能扩展用户已有偏好的圈子，不能给用户带来尝试新东西、带来新鲜感的机会。

通过账号关联，融合用户在社交网络上的特征，不仅可以推荐用户喜欢过的产品，而且还能够在新产品推荐上做到效果更好，正如 MF 方法在 k 逐渐变大时效果更好一样。矩阵分解虽然可以有效地融合多种特征数据，但不能充分考虑每一种特征最优的使用方法。

对于本章提出的融合异质社交网络特征的方法，不同于矩阵分解(MF)，将用户在社交网络上的特征划分成两类，并利用不同方法处理这两类特征，还结合了用户的历史行为，最终使用随机梯度下降方法，得到一个经实验证明有一定指导意义的结果。

小结

本章提出异质网络用户账户匹配方法(HAM)，并利用成功匹配的账户特征提高消费媒体预测的准确性。

首先，提出异质网络账户匹配的必要性以及问题定义，同时给出博弈论算法的稳定匹配思想以及计算方法。不同于普通的 link prediction 问题，稳定匹配模型假设遵循 one-to-one 的匹配，即每一个账户只能匹配一个，该假设可以有效地避免 link prediction 中长尾效应问题。

基于博弈算法稳定匹配的理解，首先针对用户在异质网络中的特征，给出异质网络账户偏好程度计算方法，其中，将相似度作为账户之间的偏好程度。扩展给出稳定匹配算法，试图匹配异质网络账户，并给出匹配算法的 3 个特性：可计算性、稳定性和唯一性，同时给出详细的理论证明其特性。

基于提出的算法，使用真实 Facebook 和 eBay 用户数据，测试算法的有效性。不论数据量大小，以及用户特征建模方式，HAM 算法性能总是高于其他算法，特别是当 $k=50$ 时，账户匹配准确度可以达到 70%，比朴素方法高 14%。除了对比计算方法，针对用户特征建模方式，给出不同建模下的匹配准确度，可以看出，基于知识库理解的用户特征建模计算结果较为稳定，不受数据量大小的影响。并且，在计算学习异质网络账户匹配后，在融合异质网络账户的预测方法中，试图利用匹配账户的社交网络信息预测目标用户在消费网络中的偏好。

参考文献

[1] Buccafurri F, *et al*. Discovering links among social networks. In: Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer-Verlag, 2012. 467-482.

[2] Carmagnola F, Cena F. User identification for cross-system personalisation. Information Sciences, 2009, 179(1): 16-32.

[3] Bartunov S, *et al*. Joint link-attribute user identity resolution in online social networks. In: Proc. of the 6th Int'l Conf. on Knowledge Discovery and Data Mining, Workshop on Social Network Mining and Analysis. ACM, 2012.

[4] Backstrom L, Leskovec J. Supervised random walks: predicting and recommending links in social networks. In: Proc. of the 4th ACM Int'l Conf. on Web Search and Data Mining. ACM, 2011.

[5] Tang J, Lou T, Kleinberg J. Inferring social ties across heterogenous networks. In: Proc. of the 5th ACM Int'l Conf. on Web Search and Data Mining. ACM, 2012.

[6] Peled O, *et al*. Entity matching in online social networks. In: Proc. of the 2013 Int'l Conf. on Social Computing (SocialCom). IEEE, 2013.

[7] Xu J, *et al*. Complex problem solving: identity matching based on social contextual information. Journal of the Association for Information Systems, 2007, 8(10): 31.

[8] Kong X, Zhang J, Yu PS. Inferring anchor links across multiple heterogeneous social networks. In: Proc. of the 22nd ACM Int'l Conf. on Information & Knowledge Management. ACM, 2013.

第四部分

辅助分析工具和案例分析

第四部分是演化计算工具和实践分析举例，包括第 12 章和第 13 章。其中，第 12 章介绍几款常用的博弈辅助分析工具 Gambit、TNGLab 和 GAMUT 后，重点介绍了用于社交演化博弈的仿真计算工具。该仿真工具针对社交演化博弈模型仿真中由随机性因素导致的需进行大量重复仿真计算的特点，将仿真任务分解成大量可独立进行仿真计算的子任务，通过分布式计算的方式来提高仿真计算的性能。该仿真工具可通过动态调整仿真子任务计算 worker 的数量来扩展仿真工具的并发处理能力，提高计算性能，为社交网络上的社交演化博弈分析提供了有力工具。

第 13 章通过对社交网络中真实数据分析的案例来说明社交演化博弈模型的应用效果。

案例 1：通过对新浪微博和 Twitter 的实际数据的采集和分析，分别估算了两个社交网络的正向行为比例、网络聚集系数和幂律指数，并与社交网络演化博弈模型仿真结果进行对比分析，发现社交演化博弈模型对于社交网络演化的仿真与真实世界中的社交网络的演化发展在一定程度上是吻合的，验证了模型的有效性。

案例 2：对于竞争性信息异步的传播，根据竞争信息影响持久度，选取了两类异步信息竞争案例：电影类和电视剧类。通过对新浪微博上热映电影和电视剧的相关微博搜索转发量的实际数据的统计分析，并结合社交演化博弈模型仿真结果进行对比分析，说明了声誉在竞争性信息异步传播中的影响。

案例 3：对于竞争性信息同步传播，选取了新浪微博中“让红包飞”活动中相关微博搜索转发的真实数据，分析相关因素对竞争传播的影响，竞争信息必须额外提供收益以弥补用户声誉损失来吸引用户改变信息选择。

案例 4：对于微博间用户使用热度的竞争性分析，选取新浪微博和腾讯微博上足球运动员账户群体，给出了其使用变化情况，并给出了分析结论。

第 11 章 演化博弈辅助分析工具

本章主要介绍适用于博弈论和演化博弈的辅助分析与计算工具，包括 Gambit[1]、Trade Network Game Laboratory[2]、GAMUT[3] 和 Flock，其中，Flock 为本书作者设计开发，用于社交演化博弈的计算分析。本书中社交演化博弈相关的模型仿真计算均使用该仿真计算工具完成。

11.1 Gambit

Gambit 是一组用于有限非合作博弈计算的软件工具包，由加州理工大学的 Richard McKelvey 在 20 世纪 80 年代中期建立。Gambit 由以下部分构成：①图形界面，用于扩展式(extensive)或策略式博弈的交互式构建与分析；②一组命令行工具，用于计算纳什均衡和其他博弈解(solution concept)；③一组用于存储的文件格式和博弈交互的工具。

Gambit 最初采用 BASIC 语言实现。1990 年左右，在 Bruce Bell 的帮助下使用 C 语言重写，并在 1991 年公开发布，版本号为 0.13。1994 年，Gambit 获得了美国国家自然基金会(NSF)的资助。在 NSF 的资助下，Gambit 使用 C＋＋重写，而通过使用 wxWidgets 库，图形界面也可以跨平台迁移。1994 年夏末，版本号为 0.94 的 Gambit 发布，1999 年 0.96 版发布，2002 年 0.97 版发布。

在 2000 年中开始，Gambit 的开发主要集中在两个方面：①图形界面开发，以遵循良好的交互设计原则为目标，尤其是对博弈论和 Gambit 的初学者具有容易的学习曲线。②Gambit 内部体系的重构，目的是使 Gambit 能够对外提供良好接口。从 Gambit 13 开始，Python API 也成为 Gambit 的一部分。

Gambit 具有如下特点：

(1) 交互式的跨平台图形界面　Gambit 的一大特点是图形界面，

它可以运行在多种操作系统下：Windows、类 Unix 系统(包括 Linux)和 Mac OS X。图形界面提供了“整合式的开发环境”，以帮助用户可视化地创建博弈模型和研究博弈模型的主要策略特点。通过图形界面，用户可以使用交互工具方便、灵活地创建扩展式博弈和策略式博弈，分析中小型博弈，获得博弈树或博弈表的可视化结果展示。图形界面适合于初学博弈论的学生使用，也可用于博弈原型的构建。

图 11-1 显示的是 Gambit 启动后的主界面，默认创建的是扩展式博弈。图 11-2 则是利用 Gambit 创建标准式博弈计算囚徒困境的纳什均衡点示例。

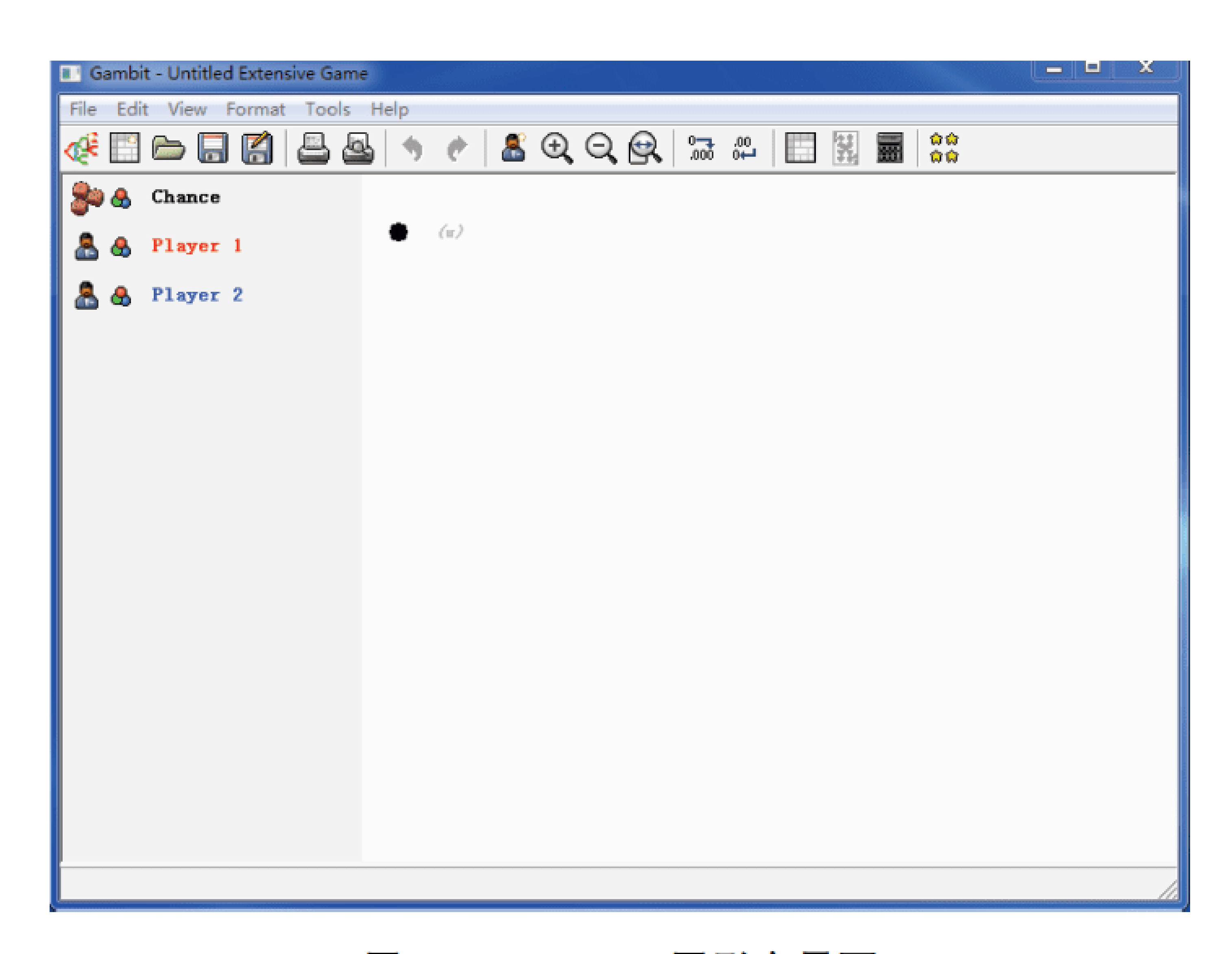


图 11-1　Gambit 图形主界面

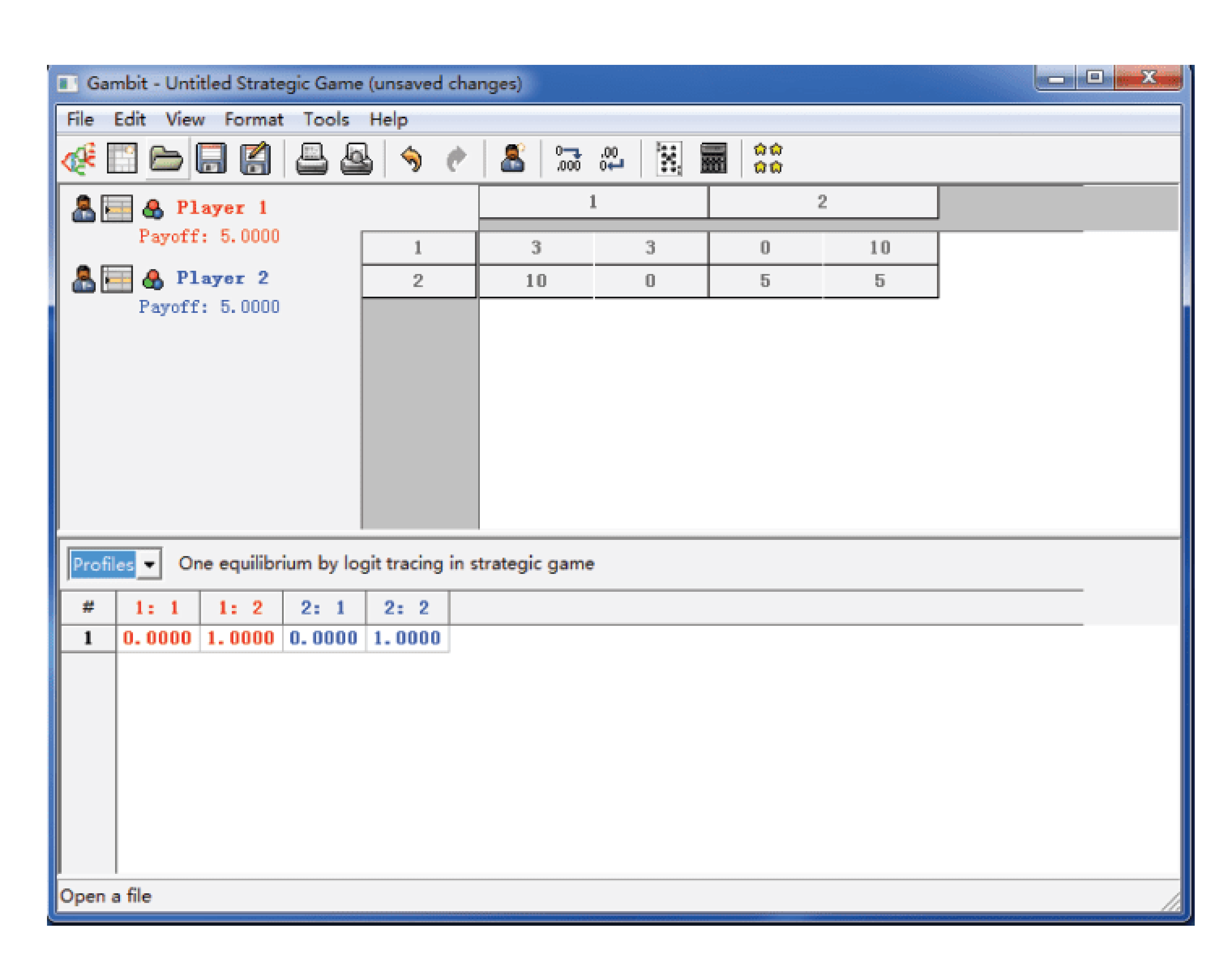


图 11-2　Gambit 图形界面计算纳什均衡点示例(囚徒困境)

(2) 用于计算均衡点的命令行工具　Gambit 的图形界面适用于分析中小型博弈，对于大型博弈分析，则需要用户直接使用 Gambit 提供的编程库。当用户需要高效计算和(或)扩展计算能力时，可以使用 Gambit 提供的命令行进行脚本运算或调用 Gambit 库进行编程。Gambit 提供的所有算法打包成独立的、命令行式程序，执行和输出都是可配置的。

(3) 扩展性和互操作性　Gambit 工具可以读写文本形式的文档化文件格式，以便于跨平台移植和外部工具的交互。这种方式可以直接扩展 Gambit 的能力，比如实现新的均衡计算方法，使用更高效的方法实现已有算法，或者程序化地实现创建、操作和变换博弈，或者基于博弈进行经济分析等。

此外，Gambit 还存在如下局限性：

(1) 只适用于有限博弈　由于有限博弈的数学结构，为这些博弈实现通用的程序是可行的。因此，Gambit 可广泛地应用于博弈应用当中。然而，当博弈不是非有限的，博弈中的局中人可能采用连续行动，或者局中人具有连续态，则无法使用通用方法。

(2) 只适用于非合作博弈　Gambit 关注于博弈分支是可以显式地写出博弈规则，并且局中人可独立选择行动。Gambit 分析工具的中心是围绕着纳什均衡，以及有限理性的相关观念，如随机最优反应均衡(quantal response equilibrium)。当前，Gambit 不提供任何表现形式或方法来分析合作博弈。

(3) 无法快速地分析大型博弈　由于一些特定的纳什均衡计算复杂度的领域仍然比较活跃，因此在一些特定案例中，计算均衡的时间复杂度会随着博弈的规模快速上升。换句话说，很容易写出需要 Gambit 运行很长时间来计算均衡点的博弈。通常，在实践中有两种方式来处理这种情形。一种方式是选用较好的启发算法来计算均衡，另一种方式是选用已知具有良好特性的博弈模型。

11.2　TNGLab

Trade Network Game Laboratory(TNGLab)是一个用于研究多市场环境下的交易网络形成的演化计算工具。该计算工具由 David McFadzean 和 Leigh Tesfatsion 等人在 2001 年开发，包含了 Trade Network Game(TNG)和 SimBioSys。TNG 是一组研究在指定市场协议规则下，交易者(买家，卖家，经销商)之间策略交互作用的贸易网络形成与演化的工具。TNG 融合和扩展了标准匹配理论和时序博弈理论，随着时间的推移，每个交易者共同决定谁进行交易(合作伙伴选择)，以及如何在交易交互中行动(策略选择)。更准确地说，在 TNG 中，资源受限交易者中的成功者基于连续更新的期望效用来选择或拒绝交易伙伴。TNG 是基于 SimBioSys 采用 C++实现的，SimBioSys 是一个用于通用演化仿真的 C++框架。

TNGLab 具有模块化和可扩展性以及清晰的易于操作的图形用户界面，可用于运行 TNG 实验以进行研究和教学。TNGLab 可以将交易网络上买家、卖家和经销商的策略交互地形成演化过程，通过网络动画可视化表示出来，并以图、表等形式来描述统计实时性能。

TNGLab 是专门针对 ACE 劳动市场研究而开发的。ACE 对经济模型进行计算研究，针对自主交互代理的演化分散系统。ACE 是专门为基本复杂的自适应系统模式设计的，主要关注经济活动过程中的宏观规律。其中，SimBioSys 是由 David McFadzean 于 1995 年基于 C++开发的框架模型，其设计目的是为了搭建具有协同演化种群的自适应交互代理的虚拟模型场景。基于 ACE 所开发的 TNG 是由 L. Tesfation 于 1995 年所开发的过程算法模型，用来研究贸易网络的形成和演化过程。

TNG 模拟贸易网络上异构买卖双方市场交互的形成与演化。

(1) TNG 过程模型在每个时间点的结构由交易双方的内部状态和行为规则决定，与系

统的需求、供给、平衡方程无关。TNG 交易行为受物理可行性限制并考虑结构一致性。TNG 中唯一的显式方程由交易双方用来描述交易场景和行为规则。

(2) TNG 中交易双方根据与他人交互行为和环境不断调整行为以满足需求。即行为规则由状态调节,并且交易双方协同调整其交互行为,表现出自组织特性。

(3) TNG 演化过程由直接作用在交易双方上的自然选择压力所决定,而不是群体水平的运动规律。这种自然选择竞争压力促使 TNG 中交易双方制定新的行为规则以保持持续开放式的交互行为。即 TNG 交易双方协同演化。

(4) 由给定初始条件出发,所有 TNG 行为活动均视初始交易互动而定,并且沿着路径依赖时间线进行交互活动。TNG 所描述的市场系统随时间变化而发展的趋势类似于培养皿中的培养生长过程。

TNGLab 计算实验室是一个具有 4 层体系结构的框架模型。框架结构如图 11-3 所示。

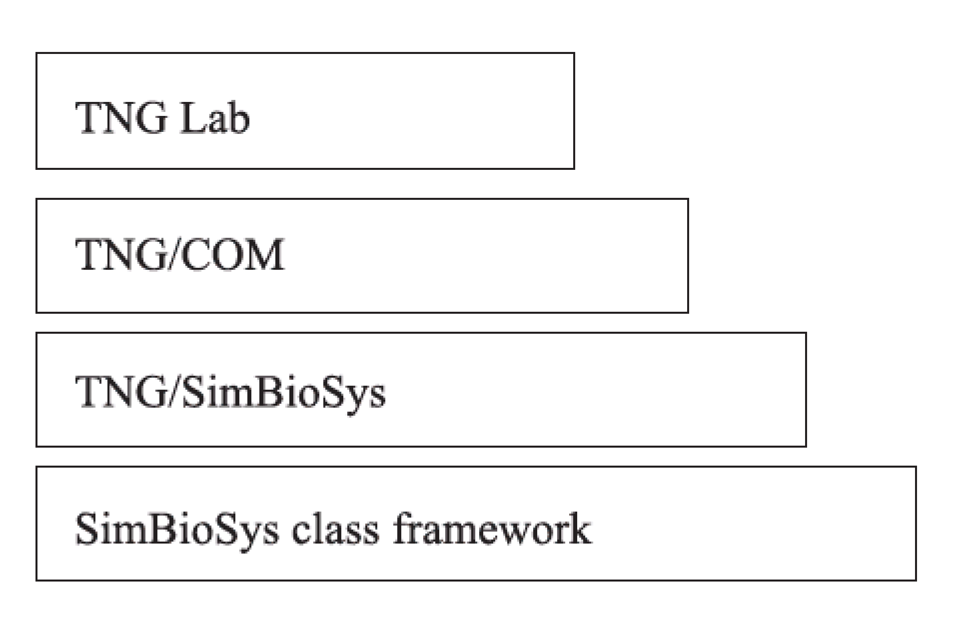


图 11-3　TNGLab 框架结构

TNGLab 最基本的底层是 SimBioSys 类框架,作为一个用于开发模拟具有自适应特性代理的群体演化的 C 语言工具包。SimBioSys 类框架实现一个设计模式来进行演化仿真模拟,控制系统的整体动态,但并不指定任何具体或者特定的行为。框架必须通过关键子类扩展来创建有效应用。

第 2 层 TNG/SimBioSys 提供扩展类,以满足贸易网络博弈(TNG)中市场协议和行为规则。TNG/SimBioSys 实现一个完整的应用程序,可以得到有价值的研究结果,但是缺乏一个可交互的友好界面。配置参数从输入文件中进行读取,并在输出文件中获得仿真结果。

第 3 层为 TNG/COM 层,通过微软的组件接口实现仿真功能,由外部程序进行通信控制。组件接口还引入一个可以实现交互式应用的事件模型。由于使用微软组件模型,所以 TNG/COM 层存在一个弊端:该框架不再是独立于该层次的平台。因为 TNGLab 针对微软 Windows 桌面系统,所以这种缺陷是可以接受的。

第 4 层 TNGLab 实现可视化图形用户界面的仿真模拟。仿真参数可以通过表单由屏幕输入,也可以从输入文件中进行保存和读取。研究人员可以手动地对可视化的应用程序进行调整编辑,并且通过一个动态显示和一个物理显示可以直观地观察可视化的仿真模拟动态。实时地仿真模拟结果通过结果显示组件以表格形式输出,同时,图表显示组件也会将同样的仿真结果数据以图像形式呈现出来。

11.3　GAMUT

GAMUT 是一个用以测试博弈理论算法的博弈生成软件。它由斯坦福大学的 Eugene Nudelman 等人在 2004 年建立。GAMUT 包含多达 35 个博弈模型类库,可以轻松实现博

弈仿真。通过设置不同参数，可以随机调节博弈分布，输入所需的常规形式博弈算法。GAMUT 博弈软件运用 Java 语言实现面向对象的框架。该框架主要包含 4 个基本类别：博弈生成器、图、函数和表达显示 4 个部分。GAMUT 软件的设计目的是为了给终端用户提供一个可以通过这 4 部分框架来描述具体的博弈类型的框架结构。

GAMUT 作为一套博弈理论测试组件，根据大量相关博弈论文献编译搭建结构化非泛型博弈综合数据库，并且建立起一个高度模块化和可扩展的软件框架来实现指定博弈模型的生成。博弈理论算法的表述性能很大程度上取决于算法运行时的实例分布，也就是说，综合表述算法数据对描述博弈算法至关重要。GAMUT 在博弈论以及计算机科学研究中将会是一个十分有用的研究工具。

GAMUT 主要包含 Games 和生成器两部分组件。Games 根据不同的非一般性博弈类型之间的关系，对博弈类型进行分类，确立不同博弈集合之间的包含关系。Games 的分类机制不仅为 GAMUT 的博弈生成器提供更深入的特征描述，更具有实际意义：即允许更为快速、直观的选择相应的博弈生成器。如果 GAMUT 并没有通过博弈生成器直接生成博弈类型，GAMUT 将生成博弈集合中随机选择集合来进行进一步的博弈生成。GAMUT 同样支持生成多个交叉集合博弈，GAMUT 均匀选择随机的生成集合。

GAMUT 的 Games 组件中包含 35 类可实现的博弈类型，均可通过程序实现仿真模拟。博弈的具体内部表述和算法依据具体博弈场景设定，能够反馈博弈中参与者的数量、每个参与者的行为次数以及每轮博弈活动给每个参与者带来的收益。Outputter 组件会将该博弈进行编码，通过适当的图表进行描述。

生成器所针对的博弈集合可以粗略分为两类。一类博弈集合可以由有效的算法程序生成，在有限时间内从集合中生成博弈样本，并且这部分集合能够生成任何类型的博弈策略。这类集合成为生成集合。另一部分非生成集合，没有合理的算法程序与之对应。通常会考虑拒绝样本方法，随机产生博弈策略，然后测试其是否属于某个特定集合。如果该集合对于大部分真实集合为非泛型集合，那么这个算法程序无法在有限时间内生成这个博弈样本，这些算法程序也不能称为博弈生成器。关系分类、博弈集合以及生成器识别是创建博弈生成器的前提。博弈算法生成器的目标是建立一个能够产生属于同一博弈集合的具有无限博弈参与者的博弈模型，并确保博弈生成器生成的博弈算法差异仅与博弈策略收益相关。通过适当的参数设置提升算法计算性能，有效地对博弈子集合实现结构化。选取能够表征特定集合的数据结构来解决算法生成中的纯算法问题。生成器组件是一个具有非平凡软件工程任务的有序易扩展的软件框架，可以轻松地实现逐步增加新的博弈算法生成器。

GAMUT 的生成器主要依赖图和函数。Graph 和 Function 两个组件以模块化形式满足生成需求。函数和图的增减更加便捷。可视化显示通过 Outputter 进行封装。为保证博弈框架的易扩展性，GAMUT 广泛支持类封装任务。GAMUT 具有强大的参数处理机制。用户可以根据指定的博弈类型、范围、默认参数值和帮助字符串创建新的生成器。用户帮助、分析以及随机化等都可以进行自动处理。通过将完整用户任务模块化，更容易实现行之有效的实际算法。其他组件提供效用促使博弈达到运算混合点以规范博弈收益。保证生成更为合理的博弈，达到纳什均衡状态，以避免机器精度等问题。

对于初始版本 GAMUT，主要针对可在计算机中存储的常规形式表现的博弈类型。这并不是限制更紧凑的博弈表现形式，例如扩展表示博弈方法，或者图论方法表示，这只是对

大型博弈案例做出限定。GAMUT 还排除了具有无限参与者数量或者行为的博弈类型以及贝叶斯博弈类型。GAMUT 并不要求博弈方法必须以规定形式进行存储，GAMUT 支持多种形式表示博弈方法，可以完整或者局部表述特定博弈类型。根据相应的博弈规则生成71 组可描述该类型博弈的程序。博弈生成集合范围涵盖从具有较小波动二乘二策略矩阵博弈到可广泛扩展参与者人数和行为数的博弈类型。所有的博弈类型均可在形象化博弈表示组件中进行紧凑编码设置。

11.4　Flock

在社交演化博弈模型中随机性占有重要的地位，如每一演化时刻中对于进入更新状态个体的选择、个体对策略更新或关系更新的选择、个体对邻居个体的选择等。为了减小模型中随机性的影响而获得更好的结果精度，在进行模型仿真时需要进行大量重复的独立仿真计算，这导致了模型仿真的计算量较大。针对模型仿真中的这一实际问题，作者设计实现了适合社交演化博弈模型仿真计算的分布式模型仿真工具来提高仿真实验的效率。本章首先介绍该仿真工具的基本架构设计，然后给出一个工作示例。本书所给出的社交演化博弈模型的仿真实验结果均是使用 Flock 仿真获得。

11.4.1　基本架构

本仿真工具包括仿真任务分解分发服务、仿真子任务计算 worker 和仿真结果服务 3 个主要部分，共分为仿真任务描述模块、仿真任务分解分发模块、仿真子任务执行模块、任务结果汇总模块和任务结果呈现模块 5 个主要功能模块。其基本功能模块架构如图 11-4 所示。

图 11-5 则说明了仿真工具部署及工作流程：

(1) 用户设置仿真参数并提交任务至仿真任务分解分发服务；

(2) 仿真任务分解分发服务收到仿真任务后，以提交任务的时间戳对仿真任务进行标记，然后把仿真任务按一次独立仿真计算作为一个子任务进行分解并构建子任务事件队列；

(3) 当某一仿真计算 worker 空闲时，它查询仿真任务分解分发服务是否有未分发的子任务。如果存在未分发子任务，则该 worker 获得一个子任务并进行仿真计算。当该 worker 完成仿真计算时，它仿真任务结果服务提交仿真计算结果，并向仿真任务分解分发服务返回子任务完成信号；

(4) 当一个仿真任务的所有子任务完成后，用户向仿真任务结果服务查询最终的仿真结果。

下面将对各主要部分及功能模块进行详细介绍。

11.4.1.1　仿真任务分解分发服务

仿真任务分解分发服务是一个服务进程，其主要功能是接受用户提交的仿真任务并对仿真任务进行子任务分解，然后将分解之后的子任务分发至仿真子任务计算 worker 进行具体的仿真计算。

在仿真任务分解分发服务中使用 libevent 和 memcached 两个开源库：

(1) libevent 是一个轻量级的基于事件驱动的高性能的开源网络库，它在仿真任务分解

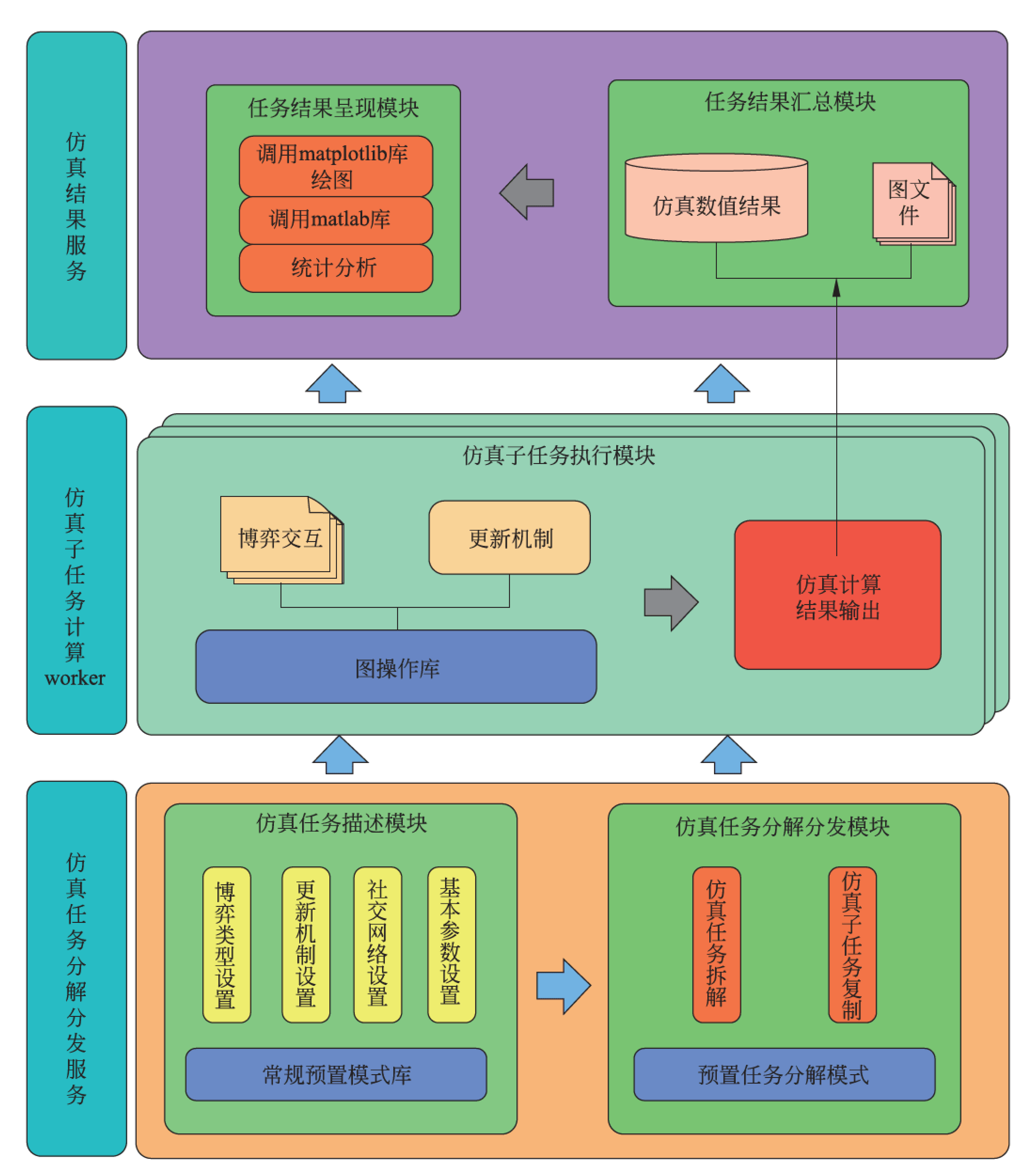


图 11-4　仿真工具功能模块框图

分发服务中主要用于处理仿真子任务计算 worker 产生的事件；

（2）memcached 是一个高性能的分布式内存对象缓存系统，它在仿真任务分解分发服务中主要用于构建仿真任务分解之后的子任务事件队列。

1. 仿真任务描述模块

仿真任务描述模块的主要功能是接收用户提交仿真任务，包括了博弈类型设置、更新机制设置、网络设置、基本参数设置以及常规预置模式库等部分。

（1）博弈类型设置：用于博弈类型的设置和博弈参数的设置，常用的博弈类型包括囚徒困境博弈、雪堆博弈、猎鹿博弈、协调博弈以及公共品博弈等；

（2）更新机制设置：用于策略更新方式的设置和关系更新方法的设置，以及相关参数

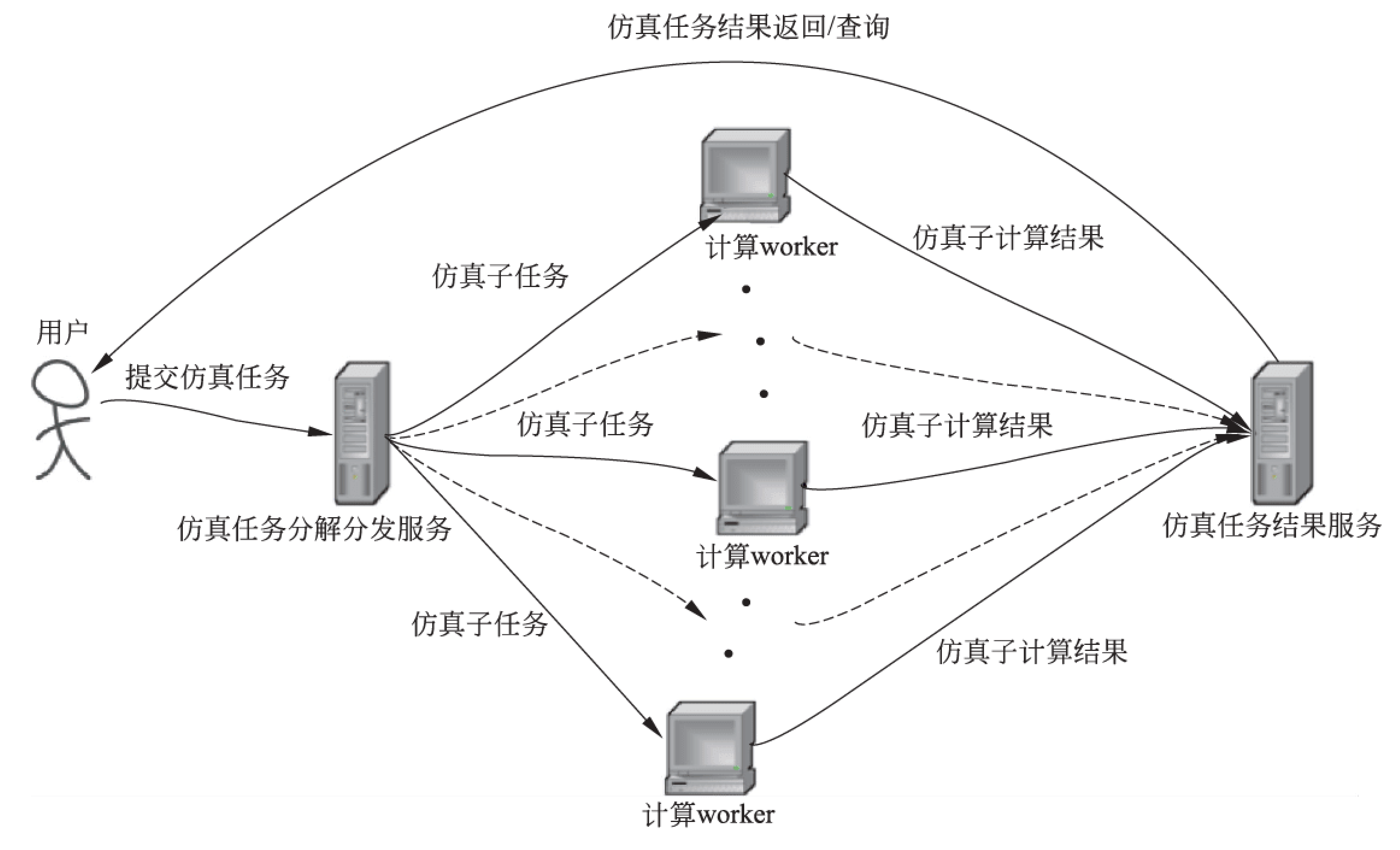


图 11-5　仿真工具部署图

的设置，常用的策略更新方式包括 Fermi 规则、更好占有规则、Best-take-over 规则等；

(3) 社交网络设置：用于社交网络生成参数的设置，如网络类型、节点数、平均度、是否有向等；

(4) 基本参数设置：用于控制仿真运算的参数的设置，如独立仿真次数、仿真周期数、采样周期数、初始合作者比例等；

(5) 预置模式库：提供了常用的仿真场景设置，如预设的网络类型、仿真周期数、独立仿真次数、默认的仿真子任务 worker 数等。

2. 仿真任务分解分发模块

经过仿真任务描述模块设置之后，仿真任务提交至仿真任务分解分发模块。仿真任务分解分发模块根据仿真任务参数的设置将仿真任务进行拆解和复制，构建仿真子任务队列。

它主要包括仿真任务拆解和仿真子任务复制两部分。

(1) 仿真任务拆解：根据参数的设置将仿真任务拆解成可独立运行的子任务。比如，当考察声誉的影响时需要对不同的p_r值进行仿真计算，在仿真任务拆解时，根据p_r的取值区间和取值步长计算出需要进行仿真的p_r值的个数，完成任务拆解；

(2) 仿真子任务复制：仿真任务拆解之后，得到可独立进行仿真的子任务后，根据重复次数对仿真子任务进行复制，构建仿真子任务队列。

3. 仿真子任务计算 worker

仿真子任务计算 worker 是进行模型仿真计算的实际执行部分，它是一个独立的进程。它启动时会向仿真任务分解分发服务进行注册，完成启动之后便可接受仿真任务分解分发

服务所分发的仿真子任务。

本部分主要由仿真子任务执行模块构成，包括博弈交互、更新机制、图操作库和仿真计算结果输出 4 部分。

（1）博弈交互：根据设置博弈类型和参数，计算个体的效用；

（2）更新机制：根据更新机制的设置，完成个体策略和（或）个体关系的更新操作；

（3）图操作库：根据社交网络设置，生成网络结构，为博弈交互和更新机制提供依据和操作空间。图操作库采用开源软件库 igraph。igraph 是一套用于复杂网络分析的开源库，提供了丰富的复杂网络计算和操作功能，并可将网络输出至多种格式，如 GraphML、Pajek 等格式。它也包含了一些网络演化博弈的策略更新算法，如 Moran 过程等。

（4）仿真计算结果输出：输出仿真计算的结果至仿真结果服务，如 f_c 随时间的采样值、终态 f_c 值、网络结构文件等需要输出的结果。

博弈交互、更新机制等算法采用动态链接库的形式提供，方便算法库的扩展。当有大量仿真子任务等待处理时，可通过动态增加仿真子任务计算 worker 运行数量的方式进行应对处理。

11.4.1.2 仿真结果服务

仿真结果服务主要提供仿真结果的汇总存储、统计分析与呈现等功能，它是由 MongoDB 数据库和一组工具包所组成。

1. 任务结果汇总模块

任务结果汇总模块提供仿真结果数据的存储和管理功能，由 MongoDB 数据库和文件存储构成。仿真结果数据主要分为两类，一类是仿真任务记录、经过大量重复独立计算的 f_c、演化过程采样数据等，存储在 MongoDB 数据库中；一类是仿真过程中输出至独立数据文件的数据，如网络结构数据等。

2. 任务结果呈现模块

任务结果呈现模块则是用于从 MongoDB 数据库和数据文件中读取数据进行统计分析、绘制网络图等功能的一组工具包。该工具包主要由 python 语言编写，通过调用 matlab 库和 matplotlib 库进行数据统计、网络图绘制等，如图 11-6 的绘制通过调用 matplotlib 库的绘图函数完成。

11.4.2 工作流程示例

本节简述模型仿真工具的工作流程。

模型仿真工具是在模型仿真实验环境中进行设置的，主要组成部分如下。

（1）运行平台：Linux 系统，版本为 Debian 7.0；

（2）使用的开源软件及开源库：curl v7.38.0，libevent v2.0-5，memcached v1.4.21，memcacheq v0.1.1，igraph v0.7.1，MongoDB v2.4.10，python v2.7.9，matplotlib v1.4.3。

本示例使用第 4 章模型中的公共品博弈进行说明，社交演化博弈模型是在有向随机网络上进行的。

仿真工具工作流程如图 11-5 所示。

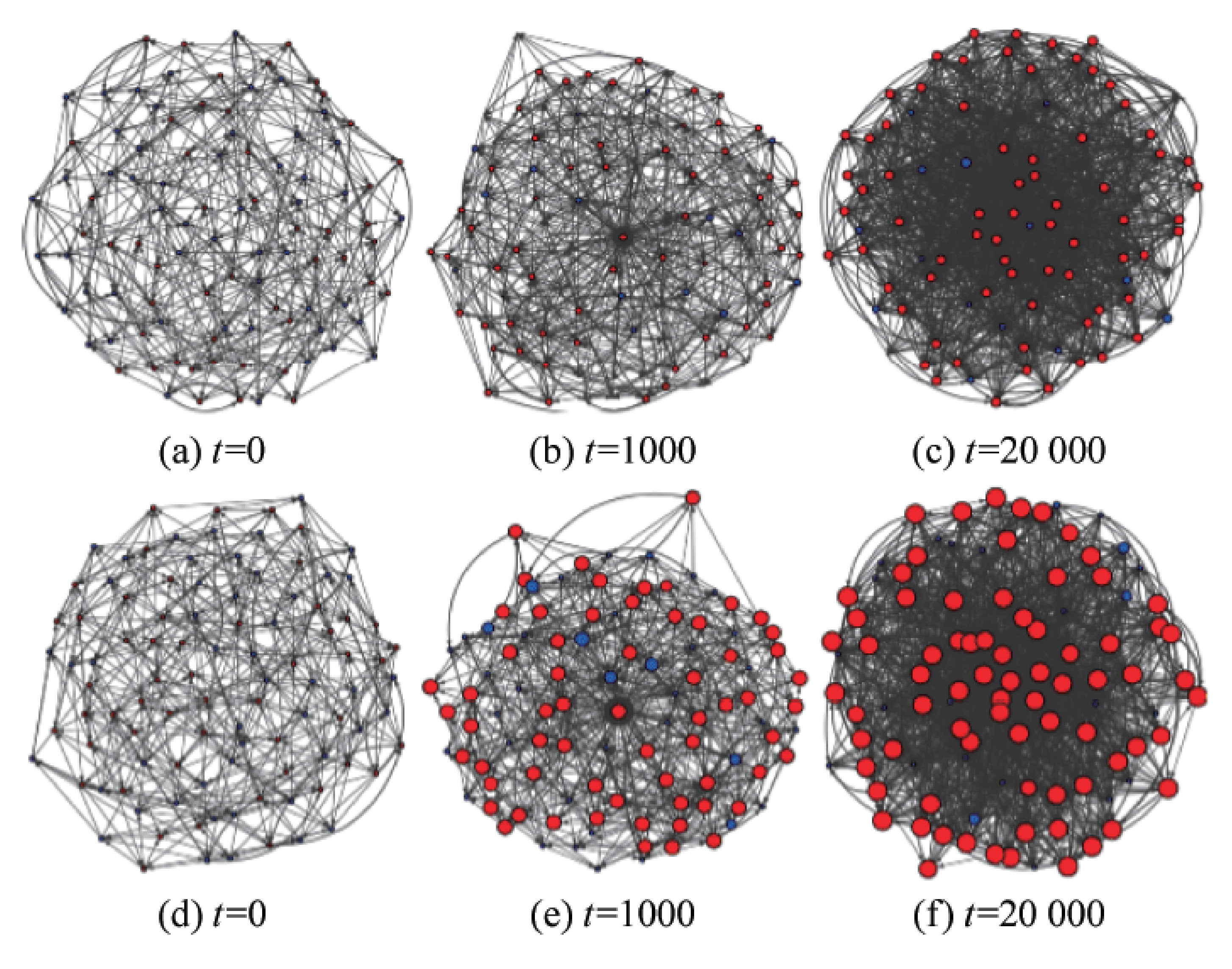


图 11-6 演化过程示例图,详见 4.3.4 节

(1) 提交仿真任务

首先,向仿真任务分解分发服务提交仿真任务。通过 Linux 终端,使用 curl 以 http 协议的 POST 方法向仿真分解分发服务监控的任务接收的端口 8669 提交仿真任务及相关参数,如图 11-7 所示。提交任务参数包括任务相关参数、网络相关参数、博弈模型相关参数、更新机制相关参数等。

```
[redacted]@debian:~[redacted]$ curl --request POST 'http://127.0.0.1:8669/'
 --data 'CMD=montecarlo&FREQ=100&PARAS={"simulation": {"timestep": 100000, "sample_times
tep":1000, "average_timestep": 1000}, "network": {"type": "random", "nvert": 1000, "k":
4, "directed": "directed", "multiple": "no_multiple"}, "game": {"type": "pgg", "r":{"max
": 3.2, "min": 1.4, "step": 0.2}, "update_mechanism": {"w": {"max": 1, "min": 0, "step":
 0.2}, "sigma": 1, "strategy": {"type": "fermi", "beta": 0.1}, "link": {"pr": 0.5, "ps":
 0.01}}, "cooper_ratio": 0.5}'
VER=0.1&CMD=transmit_data&RES=1
```

图 11-7 提交仿真任务

(2) 仿真任务分解分发

仿真任务分解分发接收到的 curl 发送的 http 协议的数据包后,首先对数据包进行解析,获得仿真任务及相关参数。然后,根据仿真任务及相关参数对仿真任务进行分解、复制等操作,构建仿真子任务事件队列。仿真子任务事件队列构建并完成设置后,处于任务空闲状态的仿真子任务计算 worker 便可接收任务转入计算运行状态。图 11-8 显示的是仿真任务分解分发服务接收到的 curl 提交的仿真任务相关调试信息。

(3) 仿真子任务计算

仿真子任务计算 worker 是一个独立计算进程,它在启动时首先向仿真任务分解分发服务的 worker 监控端口发送信息进行注册。启动完成后便处于任务空闲状态。仿真分解分发服务完成仿真子任务事件队列的构建和设置后,处于任务空闲状态的 worker 便可获得仿真计算任务进行仿真计算,如图 11-9。当仿真子任务计算 worker 完成子任务仿真计算时,它会将仿真结果发送至仿真结果服务。在本示例中仿真结果将保存在 MongoDB 中。

(4) 仿真结果存储与呈现

仿真子任务计算 worker 的仿真参数和结果存储在 MongoDB 数据库中。图 11-10 显示

```
new session
<8 server listening >
receive from ip= 127.0.0.1
new session
<10 new client connection >
accpt fd=10
line =62 rbytes=0 read_head=
line =62 rbytes=609 read_head=POST / HTTP/1.1
User-Agent: curl/7.38.0
Host: 127.0.0.1:8669
Accept: */*
Content-Length: 460
Content-Type: application/x-www-form-urlencoded

CMD=montecarlo&FREQ=100&PARAS={"simulation": {"timestep": 100000, "sample_timestep":1000
, "average_timestep": 1000}, "network": {"type": "random", "nvert": 1000, "k": 4, "direc
ted": "directed", "multiple": "no_multiple"}, "game": {"type": "pgg", "r":{"max": 3.2, "
min": 1.4, "step": 0.2}, "update_mechanism": {"w": {"max": 1, "min": 0, "step": 0.2}, "s
igma": 1, "strategy": {"type": "fermi", "beta": 0.1}, "link": {"pr": 0.5, "ps": 0.01}},
"cooper_ratio": 0.5}
rbytes=460
rbytes=460    body_len=460
body succeed
set len=460,body=CMD=montecarlo&FREQ=100&PARAS={"simulation": {"timestep": 100000, "samp
le_timestep":1000, "average_timestep": 1000}, "network": {"type": "random", "nvert": 100
0, "k": 4, "directed": "directed", "multiple": "no_multiple"}, "game": {"type": "pgg", "
r":{"max": 3.2, "min": 1.4, "step": 0.2}, "update_mechanism": {"w": {"max": 1, "min": 0,
 "step": 0.2}, "sigma": 1, "strategy": {"type": "fermi", "beta": 0.1}, "link": {"pr": 0.
5, "ps": 0.01}}, "cooper_ratio": 0.5}
set succeed
notice client
close fd=10
```

图 11-8　仿真任务分解分发

```
wait task
Get data.
{"simulation": {"timestep": 100000, "sample_timestep":1000, "average_timestep": 1000}, "netwo
rk": {"type": "random", "nvert": 1000, "k": 4, "directed": "directed", "multiple": "no_multip
le"}, "game": {"type": "pgg", "r": 1.4 }, "update_mechanism": {"w": 0.5}, "sigma": 1, "strate
gy": {"type": "fermi", "beta": 0.1}, "link": {"pr": 0.5, "ps": 0.01}}, "cooper_ratio": 0.5}
call montecarlo.
times: 0, r: 1.40, w: 0.50, pr: 0.50
start simulation.
timestamp: 1431265866499
Func games: main phase, step: 10000
Func games: main phase, step: 20000
Func games: main phase, step: 30000
Func games: main phase, step: 40000
Func games: main phase, step: 50000
Func games: main phase, step: 60000
Func games: main phase, step: 70000
Func games: main phase, step: 80000
Func games: main phase, step: 90000
Func games: main phase, step: 100000
mongo done.
simulation done.
ip = 127.0.0.1
push data.
done!
```

图 11-9　仿真子任务计算 worker 接收任务进行仿真计算

```
{ "_id" : ObjectId("554dd9556e95526a463cd621"), "timestamp" : NumberLong("1431165269534"), "t
ime_steps" : 100000, "sample_step" : 10000, "nvert" : 1000, "k_deg" : 4, "directed" : true, "
multiple" : false, "b" : 1.4, "c" : 1, "w" : 0.5, "beta" : 0.1, "pr" : 0.5, "ps" : 0.01, "sig
ma" : 1, "cooper_ratio" : 0.5 }
{ "_id" : ObjectId("554f2c556e95527c1f7c80f1"), "timestamp" : NumberLong("1431252053225"), "t
ime_steps" : 100000, "sample_step" : 10000, "nvert" : 1000, "k_deg" : 4, "directed" : true, "
multiple" : false, "b" : 1.4, "c" : 1, "w" : 0.5, "beta" : 0.1, "pr" : 0.5, "ps" : 0.01, "sig
ma" : 1, "cooper_ratio" : 0.5 }
{ "_id" : ObjectId("554f60fe6e955214a57e1791"), "timestamp" : NumberLong("1431265534170"), "t
ime_steps" : 100000, "sample_step" : 10000, "nvert" : 1000, "k_deg" : 4, "directed" : true, "
multiple" : false, "b" : 1.4, "c" : 1, "w" : 0.5, "beta" : 0.1, "pr" : 0.5, "ps" : 0.01, "sig
ma" : 1, "cooper_ratio" : 0.5 }
{ "_id" : ObjectId("554f61e26e955215110fa481"), "timestamp" : NumberLong("1431265762784"), "t
ime_steps" : 100000, "sample_step" : 10000, "nvert" : 1000, "k_deg" : 4, "directed" : true, "
multiple" : false, "b" : 1.4, "c" : 1, "w" : 0.5, "beta" : 0.1, "pr" : 0.5, "ps" : 0.01, "sig
ma" : 1, "cooper_ratio" : 0.5 }
{ "_id" : ObjectId("554f624a6e9552156d0d8e81"), "timestamp" : NumberLong("1431265866499"), "t
ime_steps" : 100000, "sample_step" : 10000, "nvert" : 1000, "k_deg" : 4, "directed" : true, "
multiple" : false, "b" : 1.4, "c" : 1, "w" : 0.5, "beta" : 0.1, "pr" : 0.5, "ps" : 0.01, "sig
ma" : 1, "cooper_ratio" : 0.5 }
>
```

图 11-10　MongDB 中存储的仿真子任务记录

的是本示例保存的部分仿真记录和结果，图 11-11 则是仿真网络的数据文件。


```
This XML file does not appear to have any style information associated with it. The document tree is shown below.

-<graphml xsi:schemaLocation="http://graphml.graphdrawing.org/xmlns http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd">
  <!-- Created by igraph -->
  -<graph id="G" edgedefault="directed">
     <node id="n0"> </node>
     <node id="n1"> </node>
     <node id="n2"> </node>
     <node id="n3"> </node>
     <node id="n4"> </node>
     <node id="n5"> </node>
     <node id="n6"> </node>
     <node id="n7"> </node>
     <node id="n8"> </node>
     <node id="n9"> </node>
     <node id="n10"> </node>
     <node id="n11"> </node>
     <node id="n12"> </node>
     <node id="n13"> </node>
     <node id="n14"> </node>
     <node id="n15"> </node>
     <node id="n16"> </node>
     <node id="n17"> </node>
     <node id="n18"> </node>
     <node id="n19"> </node>
     <node id="n20"> </node>
     <node id="n21"> </node>
     <node id="n22"> </node>
     <node id="n23"> </node>
     <node id="n24"> </node>
     <node id="n25"> </node>
     <node id="n26"> </node>
     <node id="n27"> </node>
     <node id="n28"> </node>
     <node id="n29"> </node>
Firefox automatically sends some data to Mozilla so that we can improve your experience.   Choose What I Share ×
```


图 11-11　仿真网络数据文件

小结

本章首先介绍了几款常用的博弈辅助分析工具，Gambit、TNGLab 和 GAMUT，之后，重点介绍了用于社交演化博弈的仿真计算工具。该仿真工具针对社交演化博弈模型仿真中由随机性因素导致的需进行大量重复仿真计算的特点，将仿真任务分解成大量可独立进行仿真计算的子任务，通过分布式计算的方式来提高仿真计算的性能。该仿真工具的基本架构包括仿真任务分解分发服务、仿真子任务计算 worker 和仿真结果服务三大组成部分，共包含仿真任务描述模块、仿真任务分解分发模块、仿真子任务执行模块、仿真结果汇总模块和任务结果呈现模块等。本章给出了应用该仿真工具进行前文的社交演化博弈模型仿真的示例应用来说明仿真工具工作的具体流程。

该仿真工具可通过动态调整仿真子任务计算 worker 的数量来扩展仿真工具的并发处理能力，提高计算性能，为社交网络上的社交演化博弈分析提供了有力工具。

参考文献

[1] McKelvey RD, McLennan AM, Turocy TL. Gambit: Software Tools for Game Theory, Version 15.1.0. 2014. http://www.gambit-project.org

[2] Trade Network Game Laboratory. http://www2.econ.iastate.edu/tesfatsi/tnghome.htm

[3] Gamut. http://gamut.stanford.edu

第 12 章 真实社交网络中的应用案例分析

本章聚焦于真实社交网络上的演化分析，通过给出微博上的 4 个具体分析实例，使读者对本书中前面章节的社交演化模型、方法有更进一步的理解。具体包括：Twitter 与新浪微博的信息分享行为与社交网络演化分析、竞争性信息异步传播实例分析、竞争性信息同步传播演化过程分析、用户关注度竞争分析等。

12.1 Twitter 和新浪微博的信息分享行为分析实例

以 Twitter 和新浪微博为代表的社交网络的社交功能，例如认识更多的人和维系当前的好友关系，都依赖于用户信息的分享和交互。用户分享是社交网络存在的基础。借助分享，人们表达自我、建立关系、让信息传播。如今，在社交网络上，人们看到的已经不仅仅是大众媒体，而是一个个经过个人和个人的好友们选择和加工后，分享出来的信息，形成了自媒体模式。信息的分享反映了不同人的趣味、性格和利益动机。社交网络的信息分享基于用户行为，受用户社交需求等动机因素影响，并且信息分享行为也会反作用于用户的社交关系，促使用户的社交环境发生变化。因此，利用社交演化博弈模型建模社交网络用户信息分享行为，可以分析用户信息分享行为和社交网络的演化机理。

12.1.1 Twitter 与新浪微博的演化性分析

本节通过对真实数据集的分析与模型仿真结果进行交叉验证来说明应用社交演化博弈模型分析社交网络演化机理的可行性。我们分别采集了 Twitter 和新浪微博(Sina Weibo)各约 250 000 个账号从 2011 年 10 月至 2012 年 10 月的描述和消息及转发与回复。虽然 Twitter 和 Sina Weibo 中的社交关系具有方向性，但是对消息的转发、回复、评论等具有双向性特点，并且用户进行关系调整可以增加关注或取消关注，

也可以移除他人对自己的关注。因此本文将Twitter和Sina Weibo都化简为无向网络。用户对消息的转发/回复被认为是正向行为,以下使用正向行为代称。

图12-1显示了用户粉丝数与正向行为之间的关系。在Twitter和Sina Weibo中,当用户的正向行为次数超过一定阈值(本例中是28次)后,粉丝数与正向行为呈一定的比例关系,并且越活跃的用户,粉丝数量越大。

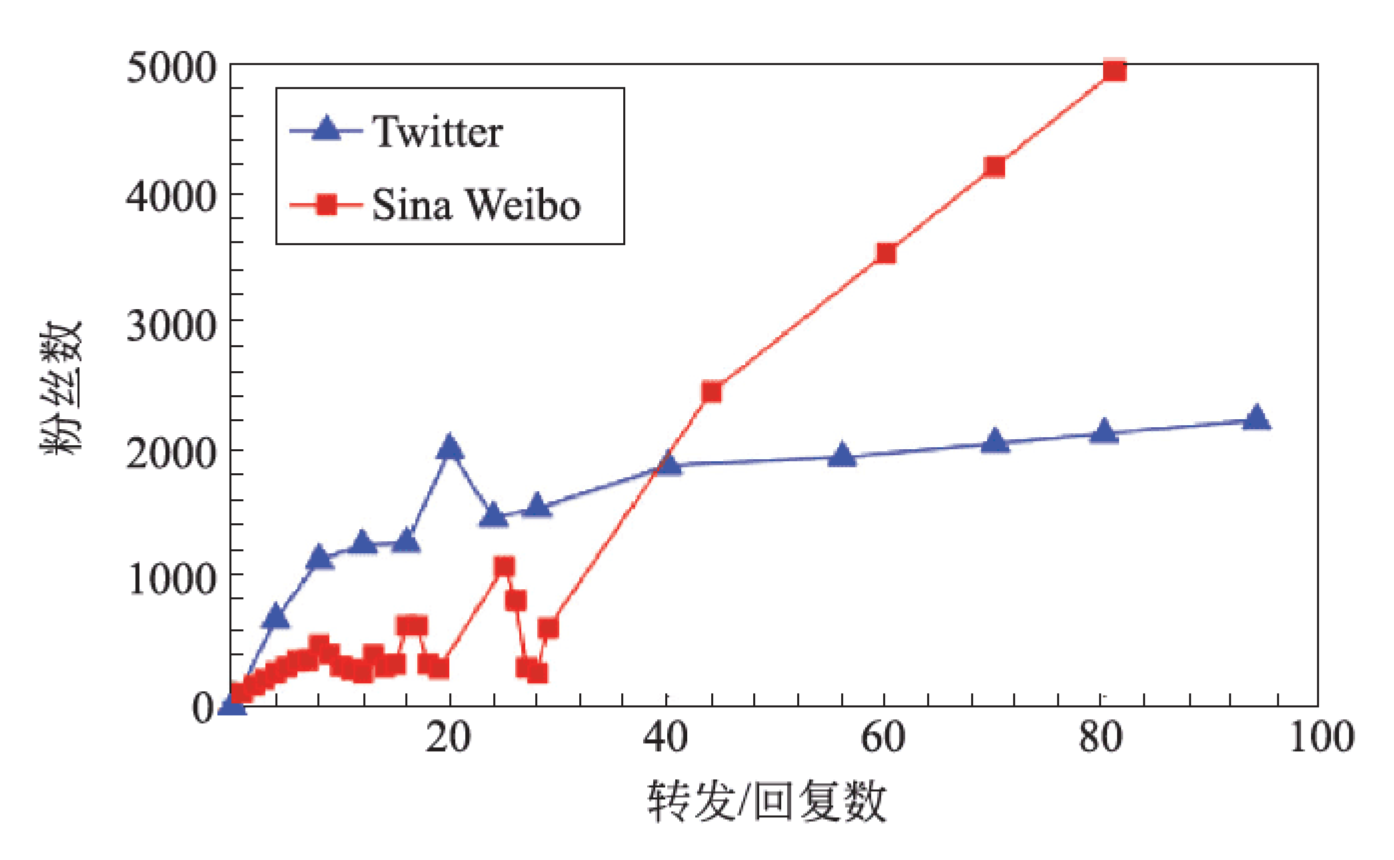


图12-1 Twitter和Sina Weibo中粉丝数与转发/回复数之间的统计分析

同时我们也看到,正向行为不足的用户与粉丝数之间的关系不够明确,可以认为普通用户受影响力、资源等因素的影响,他们比名人账号、官方账号或推广账号的行为更加难以预测。

图12-1显示了Twitter和Sina Weibo分别在2011年10月和2012年10月的粉丝度分布情况。基于图12-2的度分布情况计算了相应的幂律指数(α)。同时,我们也分别计算了用户使用正向行为的比例(f_c)和网络聚集系数(cc),结果见表12-1和表12-2。

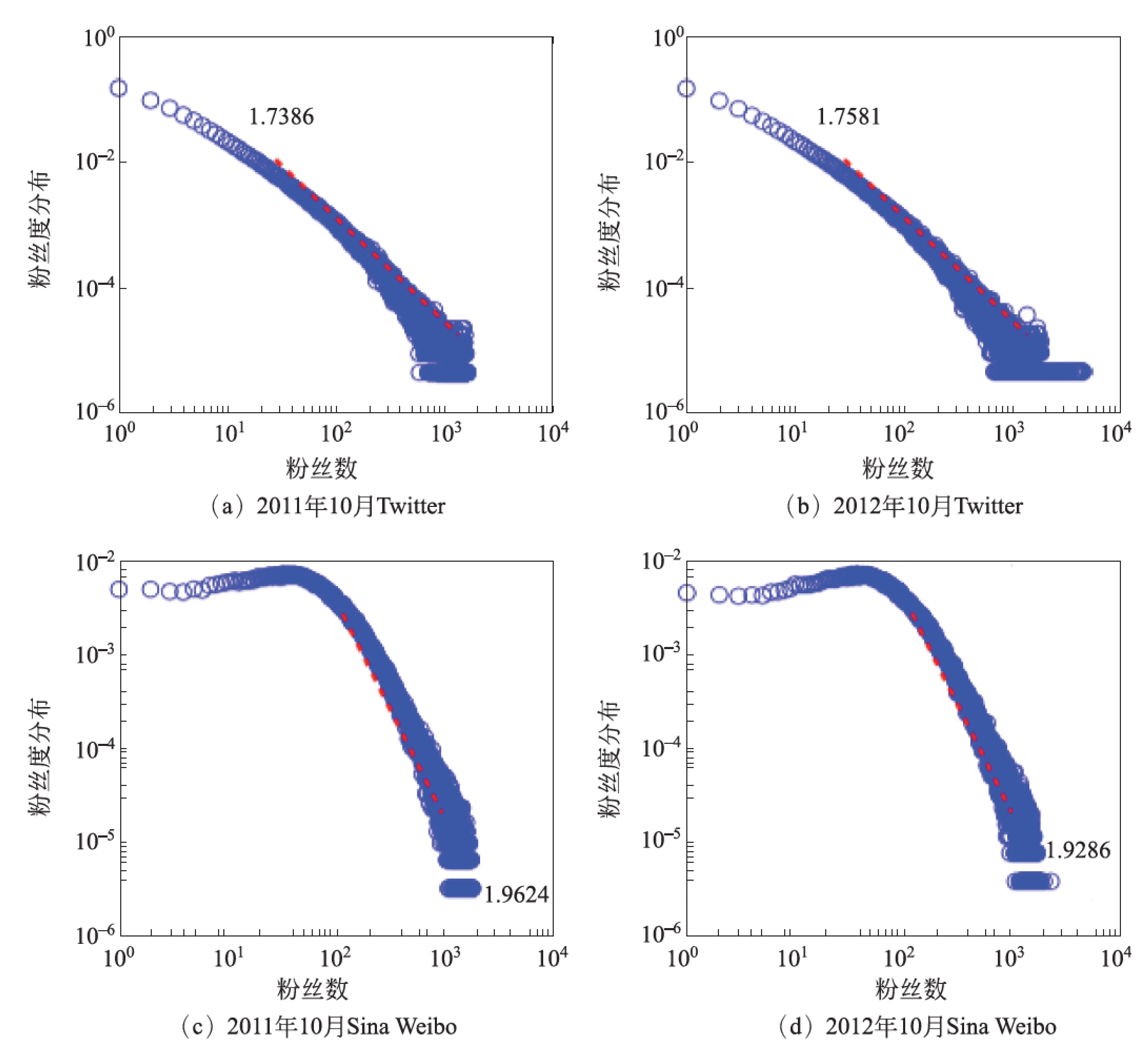
从表12-1和表12-2可以看到,两个社交网络的采样网络中2012年10月的节点数比2011年10月都减少了,原因是一些用户可能关闭了账户。相应地,Twitter采样网络的连边数也减少了,然而在Sina Weibo中则正好相反,连边数不减反增。这反映了2011年10月至2012年10月的这一年中Sina Weibo处于快速发展期,而Twitter的发展则相对平稳。对比表12-1和表12-2中的正向行为比例可以看到,Sina Weibo的用户交互活跃程度明显高于Twitter,也印证了相比Twitter,Sina Weibo处于快速发展时期。对比两个采样网络的聚集系数也显示了Sina Weibo中用户的连接关系比Twitter用户更加紧密。

12.1.2 对Twitter与新浪微博的网络特性预测

为了能够将真实数据的实证分析结果与模型仿真结果进行对比分析,我们对表12-1和表12-2中的结果分别取平均作为对Twitter和Sina Weibo的演化发展的估计值,见表12-3。

6.5.3节中图6-5的4个子图分别表示了社交网络演化的4个不同方面。这些指标对同一个社交网络来说应当对应相同的c/b和p_r。

将Twitter和Sina Weibo对应的f_c、$\overline{cc}$和$\bar{\alpha}$分别代入到图6-5(a)、图6-5(b)和图6-5(c)中得到对应的区域。如果这些指标对应的区域存在重合部分,则说明社交演化博弈模型对于社交网络演化的仿真与真实世界中社交网络的演化发展在一定程度上是吻合的,并且可



(a) 2011年10月Twitter
(b) 2012年10月Twitter
(c) 2011年10月Sina Weibo
(d) 2012年10月Sina Weibo

图 12-2　Twitter 和 Sina Weibo 粉丝度分布

表 12-1　Twitter 数据分析

时间/年	node	link	f_c	cc	α
2011	247 023	11 114 819	0.610	0.2856	1.7386
2012	246 482	11 077 792		0.2909	1.7581

表 12-2　Sina Weibo 数据分析

时间/年	node	link	f_c	cc	α
2011	257 183	35 371 927	0.935	0.3621	1.9624
2012	257 078	36 085 822		0.3811	1.9286

表 12-3　Twitter 和 Sina Weibo 数据结果平均值

	f_c	$\overline{cc}$	$\bar{\alpha}$
Twitter	0.610	0.2883	1.7484
Sina Weibo	0.935	0.3716	1.9455

以得到社交网络相应的 c/b 和 p_r。基于以上假设，可根据真实数据分析与仿真结果来验证社交演化博弈模型的有效性，并在此基础上估计 Twitter 和 Sina Weibo 的发展阶段。

把表 12-3 中 Twitter 和 Sina Weibo 的 f_c、$\overline{cc}$ 和 $\bar{\alpha}$ 分别代入图 6-5(a)、图 6-5(b)和图 6-5(c)中得到图 12-3 所示的结果。图 12-3(a)～图 12-3(c)中的紫红色区域分别表示了 Twitter 的 f_c、$\overline{cc}$ 和 $\bar{\alpha}$ 在模型仿真结果上的对应区域。相同地，图 12-3(d)～图 12-3(f)中的棕色则分别反映了 Sina Weibo 的 f_c、$\overline{cc}$ 和 $\bar{\alpha}$ 在模型仿真结果上的对应结果。把 Twitter 对应的 3 个子图的紫红色区域进行重叠之后，可以发现它们存在重叠区域，如图 12-4(a)中的红色区域所示。相同地，Sina Weibo 对应的 3 个子图的棕色区域进行重叠之后，也可以发现它们的重叠区域，如图 12-4(b)的红色区域所示。基于前面的假设和以上的结果，我们认为使用社交演化博弈在一定程度上回溯和预测 Twitter 与新浪微博的演化过程是可行的。

图 12-3(a)和图 12-3(d)分别显示的 Twitter 和 Sina Weibo 中正向行为策略比例 f_c，从这两幅图中可以看到它们的 c/b 都小于 0.2，这说明，在这两个微博类的社交网络中，用户对于他们的行为是有收益和付出方面的考虑的。此外，随着 p_r 的增大，Sina Weibo 的 f_c 区域比 Twitter 相应的区域显著增大，因此，我们推论，相比 Twitter 用户，Sina Weibo 用户更关注声誉的影响。

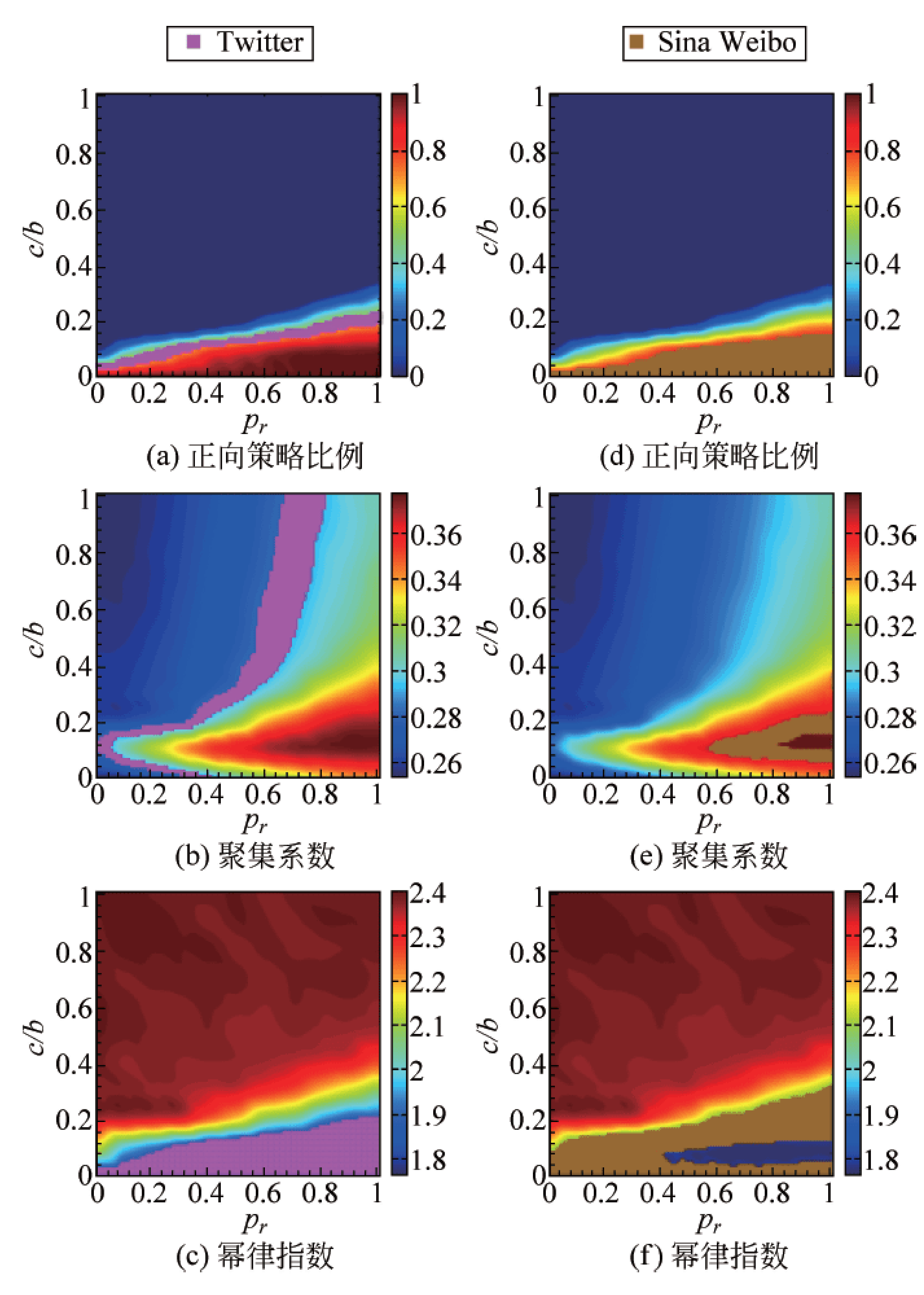


图 12-3　Twitter 和 Sina Weibo 真实数据分析结果 f_c、$\overline{cc}$ 和 $\bar{\alpha}$ 在模型仿真结果上的对比

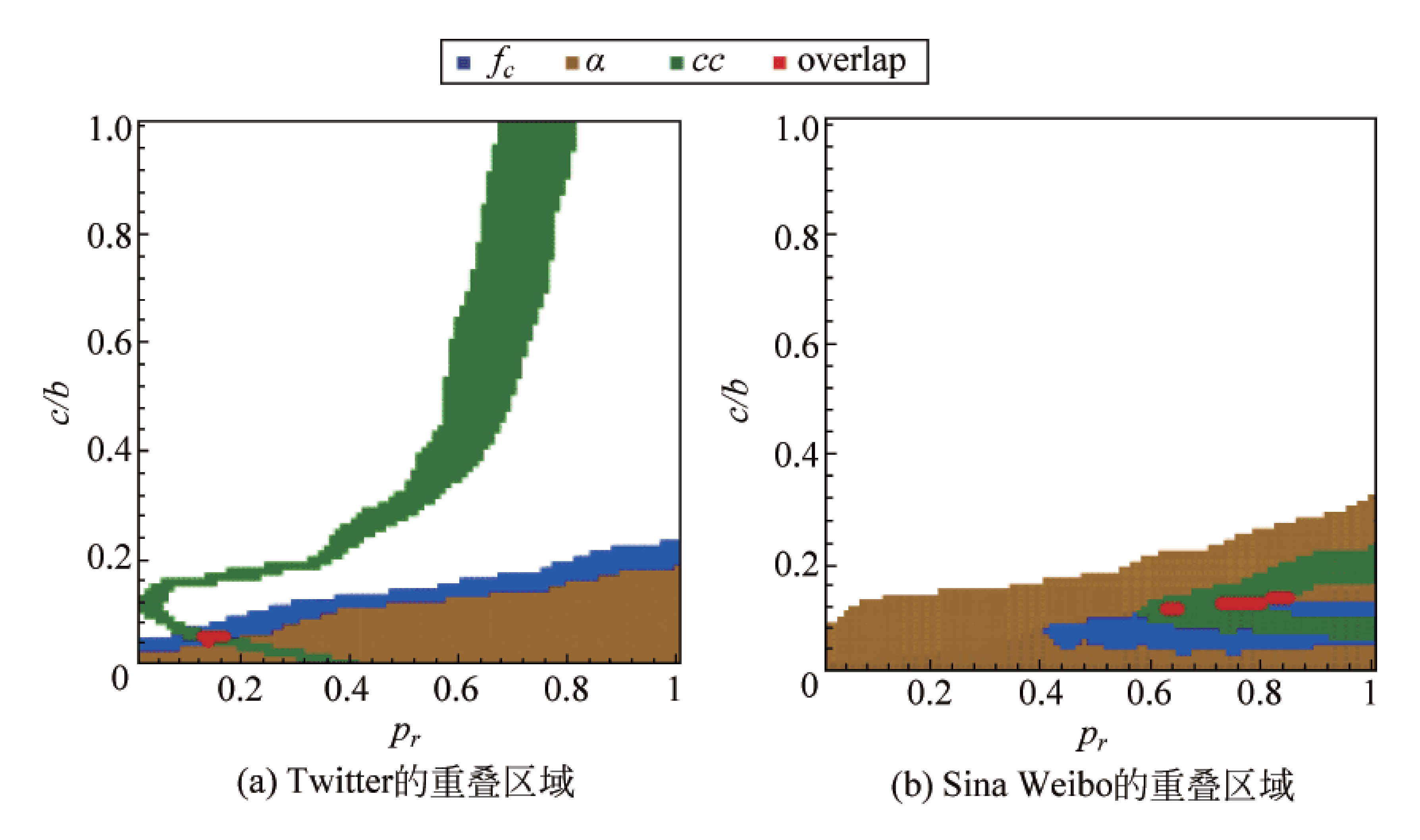


图 12-4　Twitter 和 Sina Weibo 真实数据分析和模型仿真对比结果(图 6-5)存在的重叠区域

对比图 12-4(a)和图 12-4(b)中的重叠部分可以观察到，Sina Weibo 的用户相比 Twitter 用户更愿意为正向行为(如转发、评论等)而承担更多的成本，并且更多地关注他们的声誉。通过图 12-4 中的两个子图，我们可以分别得到 Twitter 和 Sina Weibo 对应的$\overline{c/b}$和 $\overline{p_r}$，如下所示：

- Twitter：　$\overline{c/b} \in [0.04, 0.05]$，　$\overline{p_r} \in [0.13, 0.16]$
- Sina Weibo：　$\overline{c/b} \in [0.11, 0.13]$，　$\overline{p_r} \in [0.62, 0.84]$

将两个社交网络得到的重叠部分的$\overline{c/b}$和$\overline{p_r}$对应的区间代入图 6-5(d)，得到图 12-5，从而得到 Twitter 和 Sina Weibo 演化时间的估计。基于图 12-5，我们分别计算了 Twitter 和 Sina Weibo 的平均演化时间t_{Twitter}和$t_{\text{SinaWeibo}}$，两者的比值$\frac{t_{\text{Twitter}}}{t_{\text{SinaWeibo}}}=1.0373$ 显示出，相比 Twitter，Sina Weibo 需要更多的时间演化到相对稳定的状态。

通过对社交演变博弈模型的仿真和 Twitter 与 Sina Weibo 真实数据集分析的对比，验

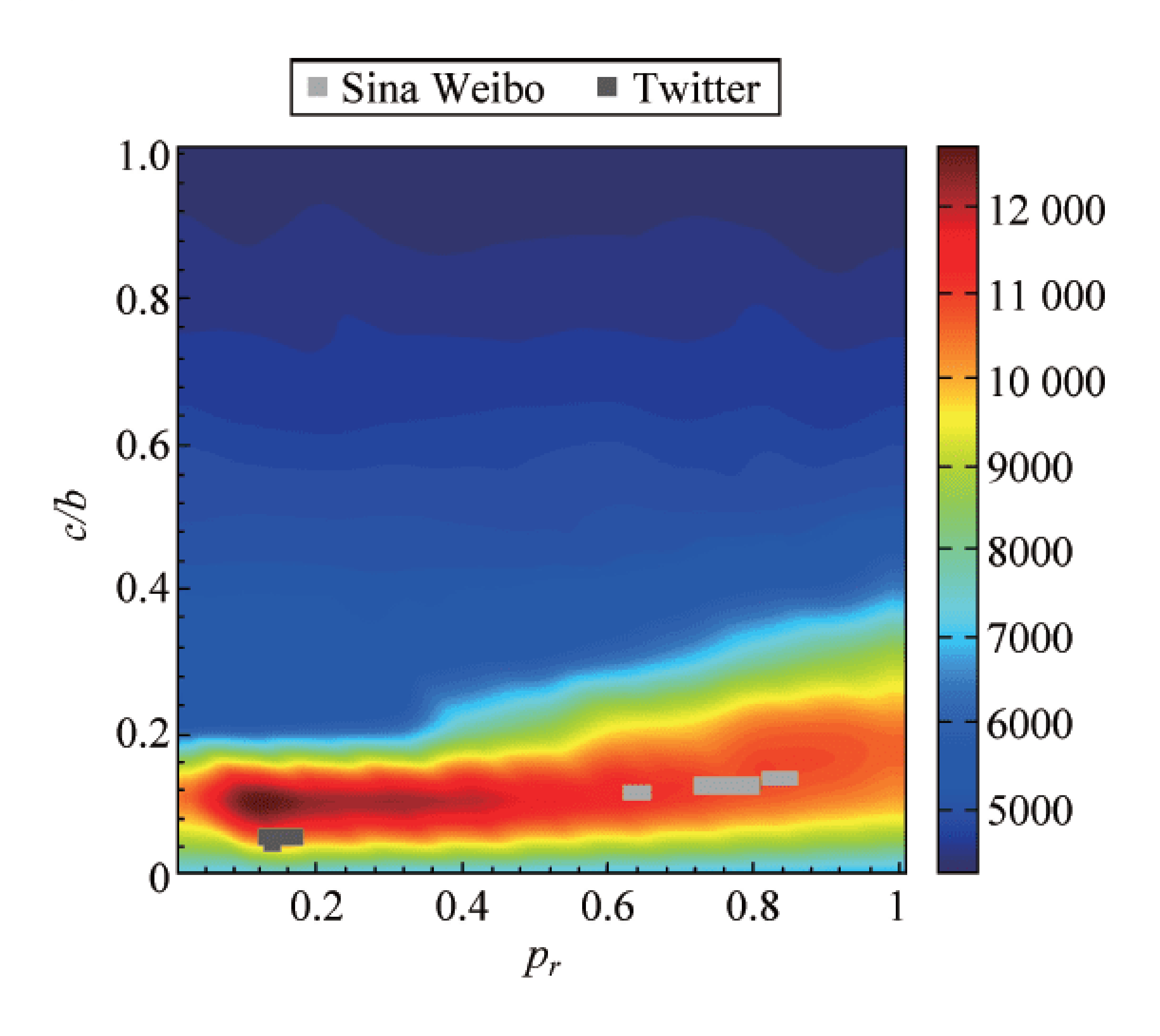


图 12-5　演化时间估计

证了使用社交演化博弈模型研究社交网络演化的可行性以及 Twitter 和 Sina Weibo 的演化。模型的仿真结果显示，基于声誉的社交关系更新能够使网络中的正向行为保持稳定并引导社交网络达到相对稳定的状态。

在社交关系更新和策略更新的共同演化中，社交网络中的社交关系的星形结构以很大的概率加速了社交网络的演化，而其中，声誉在社交关系演化中扮演了重要的角色。进一步地，通过对采集自 Twitter 和 Sina Weibo 各 1 年的数据进行分析后，我们对比了数据分析结果与模型仿真结果。正向行为促进了社交关系的形成而声誉则会加强这种影响。相比 Twitter 用户，Sina Weibo 用户能够承受更大的付出，进行转发/回复等正向行为，同时，Sina Weibo 用户比 Twitter 用户更关注声誉。

12.2　新浪微博中电影和电视剧的竞争传播分析实例

电视剧和电影传播媒介与过程的差异导致两类竞争信息在网络上呈现的竞争情形存在区别。电视剧最鲜明的传播特征之一在于家庭式的日常接收环境。每个频道都是一个信息流，每天不间断地滚动播出，而电视剧就在这多个电视流背景中，在固定时间单向传播。观众往往有意或无意地从这些电视流中选择电视剧来观看。因此，故事的连续性、播出时间是否合理、是否有观众喜爱的明星参演、是否具有知名度、类型是否合乎大众喜好等因素决定有多少观众会花费多长时间停留在一部特定的电视剧上面。电影相对电视剧长时间的更新播出，更多地是将一个故事在短时间内交代清楚，上映时间也更为短暂，一般从上映到下映为 1 个月左右的时间。因此电影的前期准备、营销推广更被看重。影片的上映更能在短时间内在观众人群中产生强烈的讨论。

电影往往以强调“质量”来提升票房口碑，而电视剧更依赖于连续的“数量”来赢得观众关注。电影通常 100 分钟左右叙述故事，而电视剧通常在数十天乃至几个月中连续播出。电视剧与电影的最大不同在于其连续性与片段性相结合，这一特点表现在社交网络的信息传播上，电影相关信息更多地依靠影片前期宣传作为铺垫，上映短期内口碑推荐，相关信息往往会呈现短期内极具爆发式的增长或者下降，网络用户对影片的口碑往往对影片相关信息的传播起到至关重要的作用。电视剧在上映前往往相关信息并不会引起网络用户的重视，而是随着电视剧的播出，逐渐在网络产生一定的影响。某一集剧情的突出，或者明星的加盟会在一定程度上刺激信息的传播。在整个电视剧的播出阶段，信息的竞争更是一个相对长期博弈的过程，相对电影的短期竞争并没有那么剧烈。

12.2.1　新浪微博中电影竞争信息异步传播分析

本节我们选取了在国内极具影响力的在线社交网络“新浪微博”的电影数据进行竞争信息异步传播的分析。为了保证模型分析验证的准确性和可靠性，我们选择目前最具影响力之一的在线社交平台——新浪微博作为实验平台。微博平台可以视为一个具有较低声誉因素，网络个体双向沟通交流的有向网络。我们采集了 2015 年 7 月新浪微博上 700 000 多个用户关于两部热映电影相关微博的搜索和转发数据。在 7 月同档期先后上映了两部热门影片《小时代 4》和《大圣归来》，两部电影在微博上的扩散讨论可视为两类竞争性信息在社交网络上的扩散传播，先上映的电影《小时代 4》视为微博观影人群中原有的竞争信息 A，而后

上映的《大圣归来》视为后加入竞争的信息 B。我们抽取两部电影相关的热搜度和热议度，分别用来描述两部电影对观影人群的吸引力和相关电影信息在观众中的传播范围及其影响力的大小，对相关数据处理后分别对应于模型仿真实验中的两个竞争信息的收益效用 a 和 b，以及演化博弈中的平均合作率 f_c。微博观影人群具有一定的网络焦点效应：观影人群的注意力趋向于集中到同一关注的信息资源上，偏向于共同推荐转发某一电影相关资讯。这一特征很好地符合了本文所采用的协调博弈策略，能够很好地描述社交网络上这一现象。

我们最终获得新浪微博上移动端对"小时代 4"和"大圣归来"两个关键词的搜索量和《小时代 4：灵魂尽头》和《西游记之大圣归来》两部电影在微博上的热议度，即两部电影关键词在微博平台的提及次数，包括相关博文的原创和转发数量。搜索量和热议度作为两部电影相关信息的效用值和传播程度。基于关键词每日的微博热议度，以关键词为统计对象，计算出各个关键词在新浪微博中的热议趋势，并以曲线图的形式进行展现。通过热议度来反映微博相关信息的发展走势。采集的数据全部来源于新浪微博数据中心，并对垃圾账号和垃圾网站予以过滤，保证了各项数据的公正性。

通过对新浪微博中《西游记之大圣归来》和《小时代 4：灵魂尽头》两部电影的传播过程及特性进行分析，对比 5.2.8.1 节的模拟实验结果，对两部电影作为竞争信息在新浪微博上的传播演化过程进行讨论，验证了模拟仿真的可行性，并对社交网络竞争性信息的传播给出相关建议。

1. 热议度分析

我们对新浪微博上的用户在 2015 年 7 月 1 日至 7 月 31 日的一个月内对《西游记之大圣归来》和《小时代 4：灵魂尽头》的热议程度的相关数据进行采集分析。统计相关词条"大圣归来"和"小时代 4"原创和转发的博文数，描述出两类关键词被提及的变化趋势。两类关键词的热议程度表现出《西游记之大圣归来》和《小时代 4：灵魂尽头》两部电影相关信息在新浪微博这一社交平台上的扩散范围以及观影群众中的影响程度，而仿真实验中的平均合作率 f_c 可以用来描述这一特性。f_c 描述群体中原信息 A 在群体中所在比率，即表现为在观影群体中《小时代 4：灵魂尽头》相关信息的热议度所占百分率。根据电影上映时间以及热议趋势可以将整个趋势划分为 4 个阶段：准备期、宣传期、成熟期和持续期。以《西游记之大圣归来》为例，4 个阶段如图 12-6 所示。《西游记之大圣归来》首映礼为 7 月 7 日，之前的时间段，该电影一直以较低姿态在微博中扩散，整个准备期中并没有在微博平台上掀起强烈讨论。首映礼之后到 7 月 10 日正式上映的宣传期内，良好的影片质量以及营销推广使得《西游记之大圣归来》在微博中赢得广大网友的关注。影片上映之后，微博媒体的持续造势，观影人群的积极追捧，使得《西游记之大圣归来》迅速成为公共舆论中的重要话题，热议程度得到井喷式的发展。票房更是在 13 日破亿，该日的热议度也达到峰值，选取 7 月 15 日票房达 2 亿作为成熟期与持续期的分界点，7 月 15 日之后热议程度开始逐渐下降，趋于平缓。

图 12-7 中表示出两部电影的热议度的变化趋势。从图中可以看出，两部电影的热议度均呈现出先上升后下降的波动趋势。从该图中可以很明显地看出，7 月初的微博平台上《小时代 4：灵魂尽头》属于重点讨论影片，《小时代 4：灵魂尽头》由于先上映，在微博平台相关信息竞争传播中抢得先机，而《西游记之大圣归来》尚未引起大众关注。随着《西游记之大圣归来》的上映，由于电影良好的制作，微博上成功的推广营销，大众影迷自发的宣传介绍，使

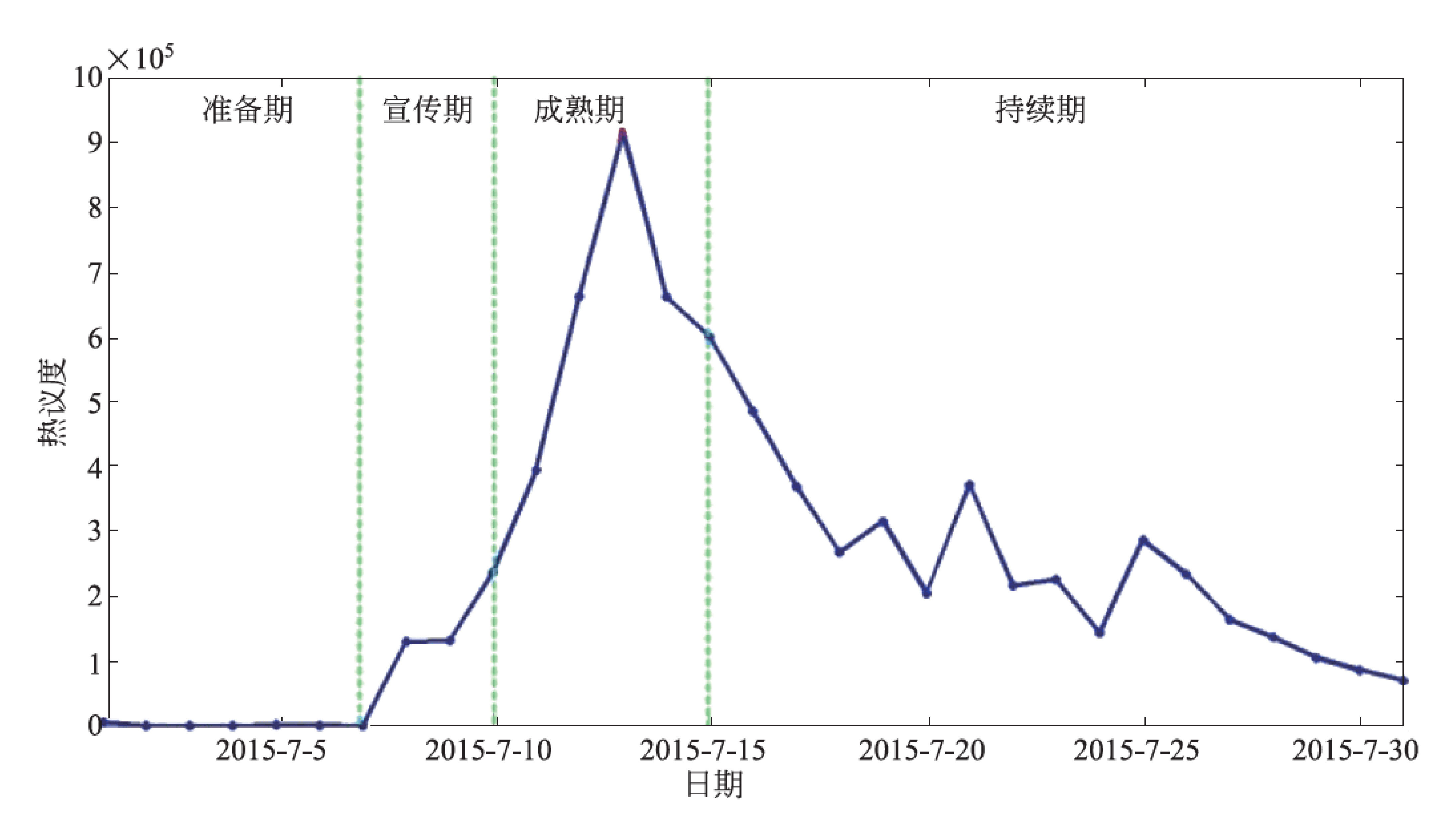


图 12-6　《西游记之大圣归来》热议度日趋势

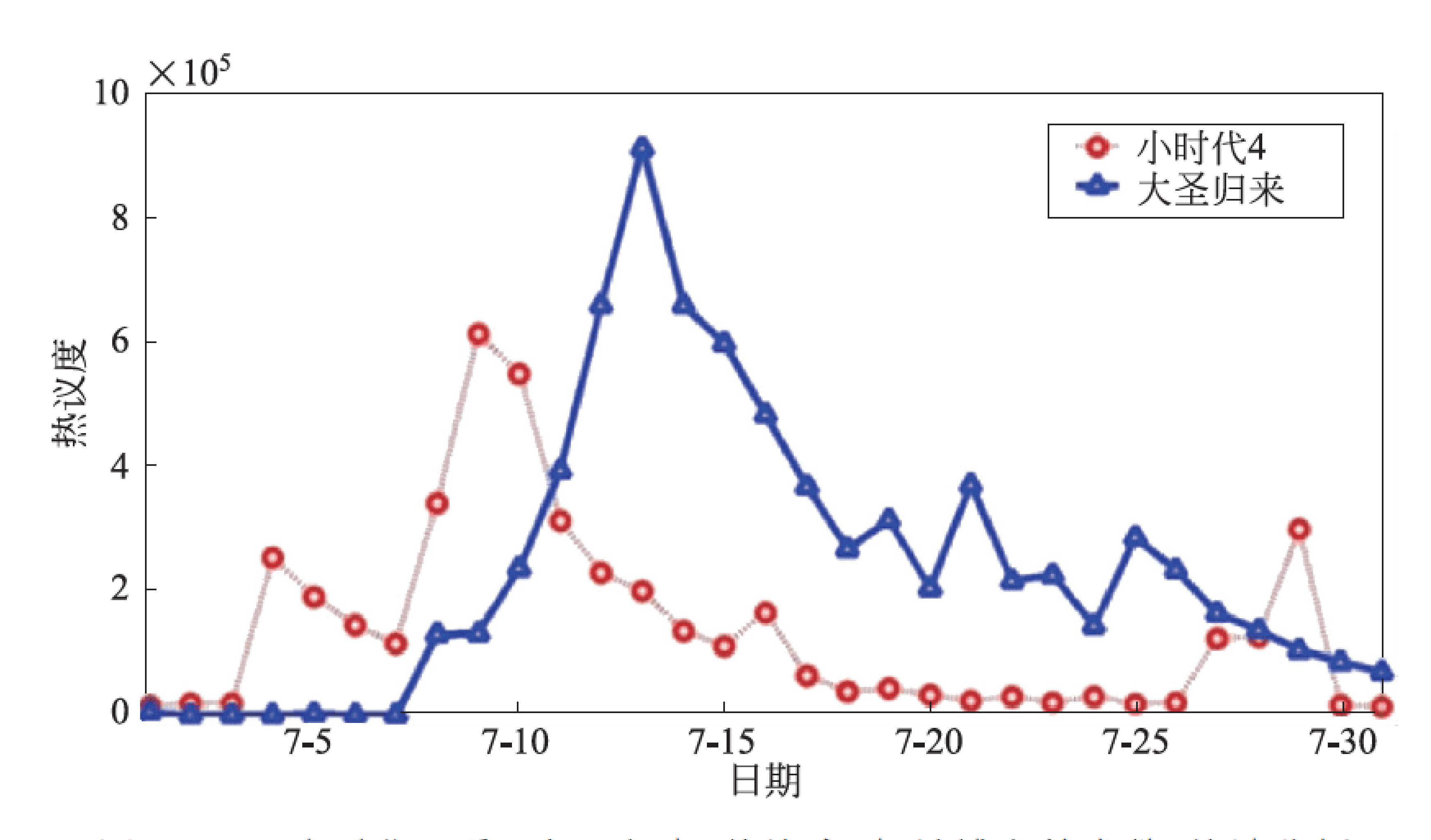


图 12-7　《小时代 4》和《大圣归来》热议度(相关博文转发数)统计分析

得相关话题在微博上迅速传播开来，成为微博上网友议论的焦点。在热议度上呈现出《西游记之大圣归来》对《小时代 4：灵魂尽头》压倒性的优势。随着影片的持续热映，相关话题讨论也逐渐下降，但《小时代 4：灵魂尽头》仍然处于劣势。在 7 月末，《小时代 4：灵魂尽头》有一个短时间段的热议度上升占据竞争优势的阶段，这一现象在之后也会详细讨论。

图 12-8 中平均合作率f_c 的值表示在与《西游记之大圣归来》的竞争中，《小时代 4：灵魂尽头》的相关话题在微博平台上的所占比例。由f_c 的变化情况来反映《西游记之大圣归来》和《小时代 4：灵魂尽头》在新浪微博上的相关信息竞争传播的演化过程。红色虚线是 0.5 的f_c 值的界限，蓝色折线是f_c 值的变化情况，当f_c 值大于 0.5，折线位于虚线上方时，说明用户偏向原有信息 A，更倾向于坚持原有策略，《小时代 4：灵魂尽头》在竞争中更受欢迎；当f_c 小于 0.5，折线位于虚线下方时，说明用户偏向新加入的信息 B，倾向于抛弃原有信息转而接纳新信息，《西游记之大圣归来》更能对用户产生影响力。f_c 值越高，《小时代 4：灵魂

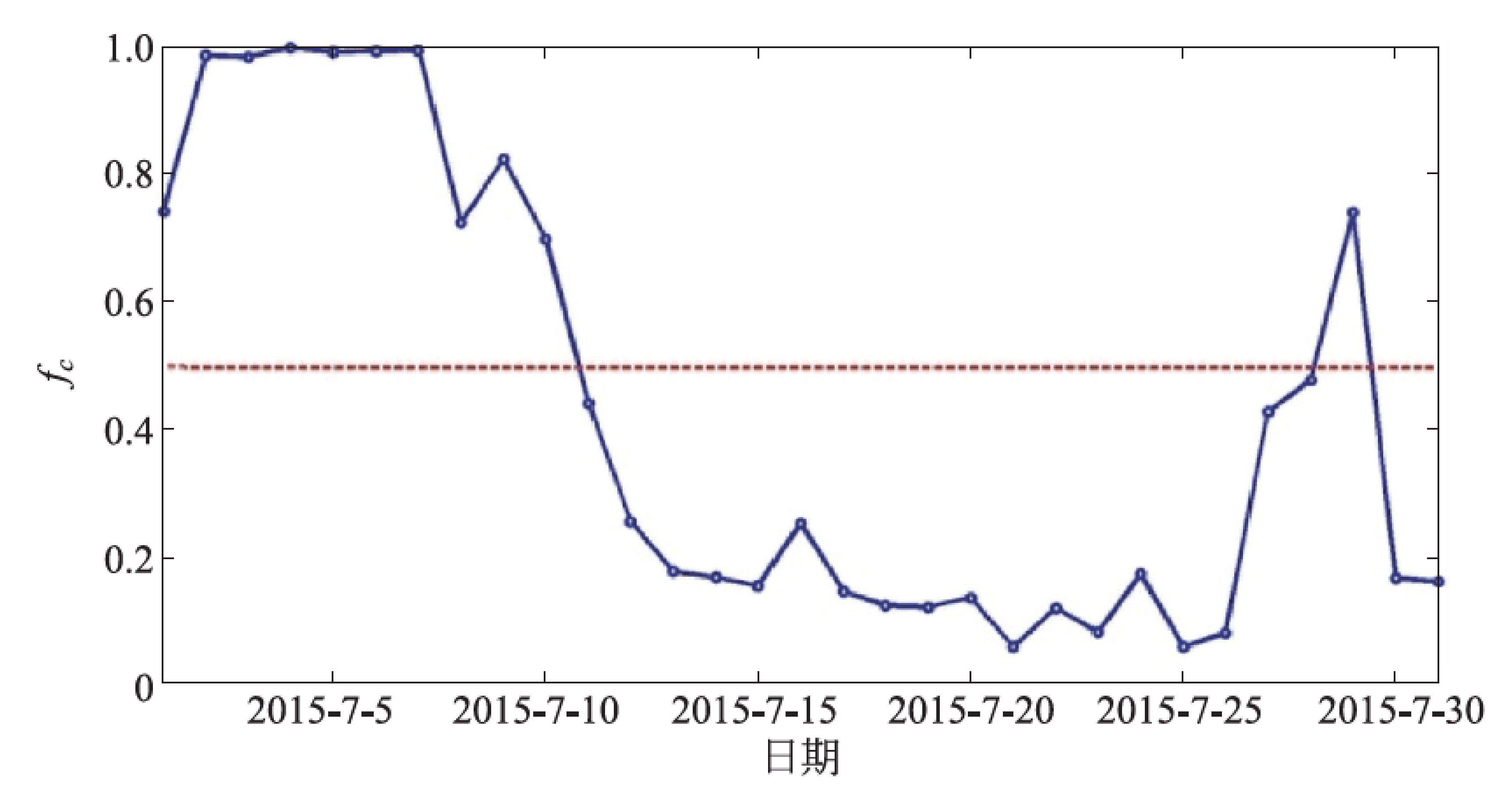


图 12-8　合作率 f_c 变化趋势

尽头》相关信息在微博平台上的讨论量越大，影片越受关注；反之，《西游记之大圣归来》在竞争传播中更具优势。从图中可以看出，在 7 月上旬，抢先上映的《小时代 4：灵魂尽头》相关消息的传播占据绝对优势，但是随着《西游记之大圣归来》的上映，《西游记之大圣归来》的相关信息迅速在网络上传播，在信息传播的竞争中一举扭转劣势，在信息竞争传播中占据主动。可以视为后加入竞争的信息凭借自身具有的较高效益吸引了网络个体接纳，在竞争传播中成功打败网络中原有的信息。其中一个有趣的现象值得关注，在 7 月末，《小时代 4：灵魂尽头》信息占有率出现一个峰值，通过相关话题在微博上发起讨论并引发关注，在与《西游记之大圣归来》的竞争中又短时间地占据一定优势，在这个阶段内，竞争关系出现反转，《西游记之大圣归来》成为原有的信息，而《小时代 4：灵魂尽头》成为后加入重新竞争的新信息，成功在与原信息《西游记之大圣归来》的信息竞争传播中占据一定优势。

2. 吸引力分析

我们采集“大圣归来”和“小时代 4”相关词条在微博客户端的搜索量，以此作为两部电影对网络用户的吸引力效用。搜索量越大，表明该电影对用户的吸引力越大，该信息在博弈演化中的效用越高。将原有竞争信息“小时代 4”的效用作为标准 1，新加入竞争的“大圣归来”效用值以“小时代”效用值为标准，作为一个变量 b 相应波动变化。

图 12-9 描述了在 7 月内两部电影对观众吸引力效用的变化趋势。红色虚线表示“小时代 4”相关话题效用标准 1，相应地，蓝色折线反映了“大圣归来”相关话题效用的变化。7 月 10 日之前，蓝色折线都在红色虚线之下，说明“小时代 4”效用更具优势，搜索量更大。7 月 11 日开始，蓝色折线波动到红色虚线之上，“大圣归来”的效用开始增大并反超“小时代 4”，用户在微博平台上对“大圣归来”的搜索量大大增加。7 月 28 日和 29 日，“小时代 4”在微博上掀起一波话题讨论，成功地增大了用户对其的关注度，搜索量随之上升，效用值也一度反超“大圣归来”。

3. 对比分析

我们将 f_c 随 b 变化情况的波动趋势用折线图绘制出来，发现与图 5-19 中 p_r 值较低时的演化特性较为符合。我们单独选取 $p_r=0.1$，$W=0.2$ 条件下演化过程中 b 和 f_c 的变化值，

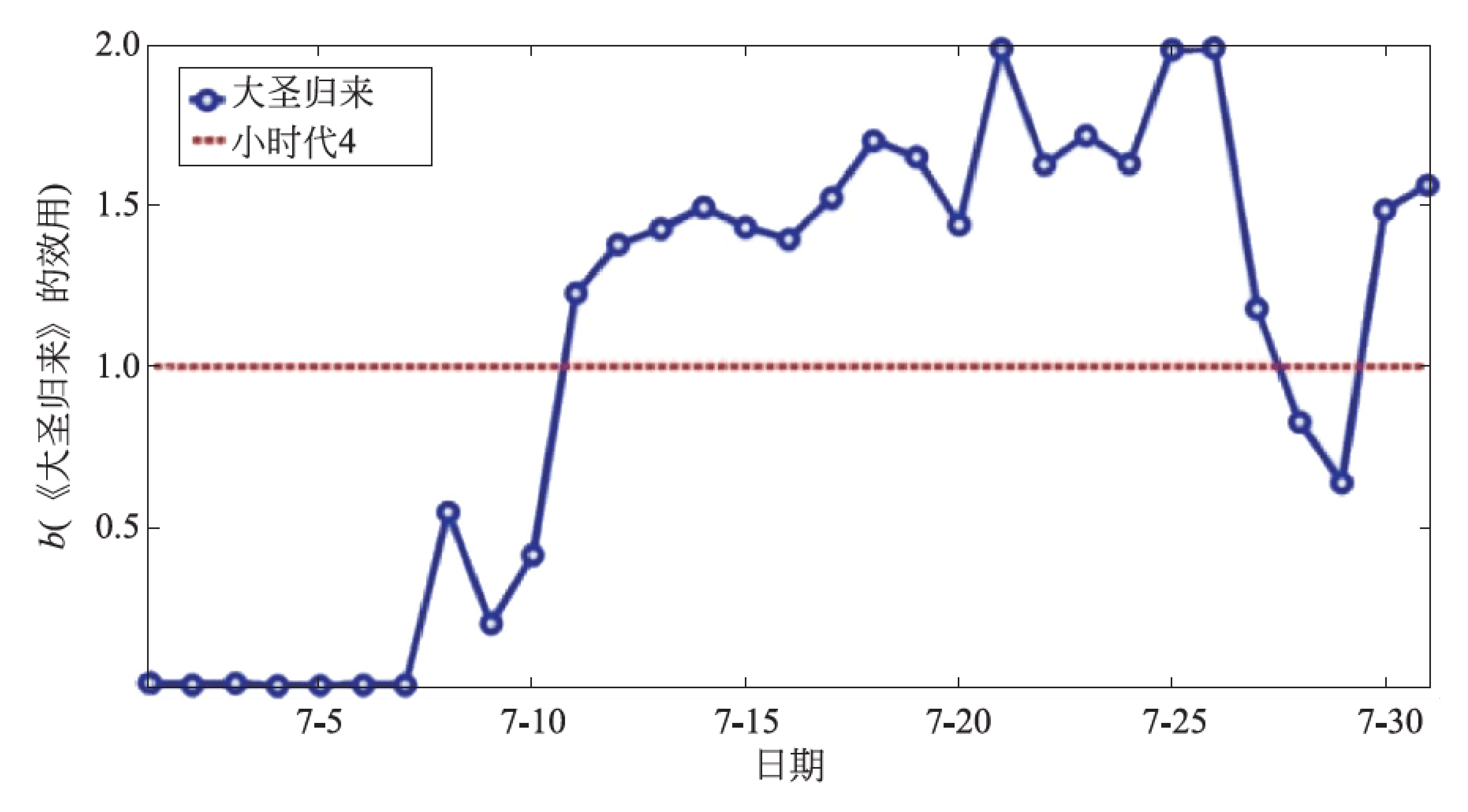


图 12-9 《小时代 4》和《大圣归来》效用(相关关键词搜索评论数)统计分析

将仿真变化曲线与实际变化曲线绘制在一张图中进行比较,如图 12-10 所示。

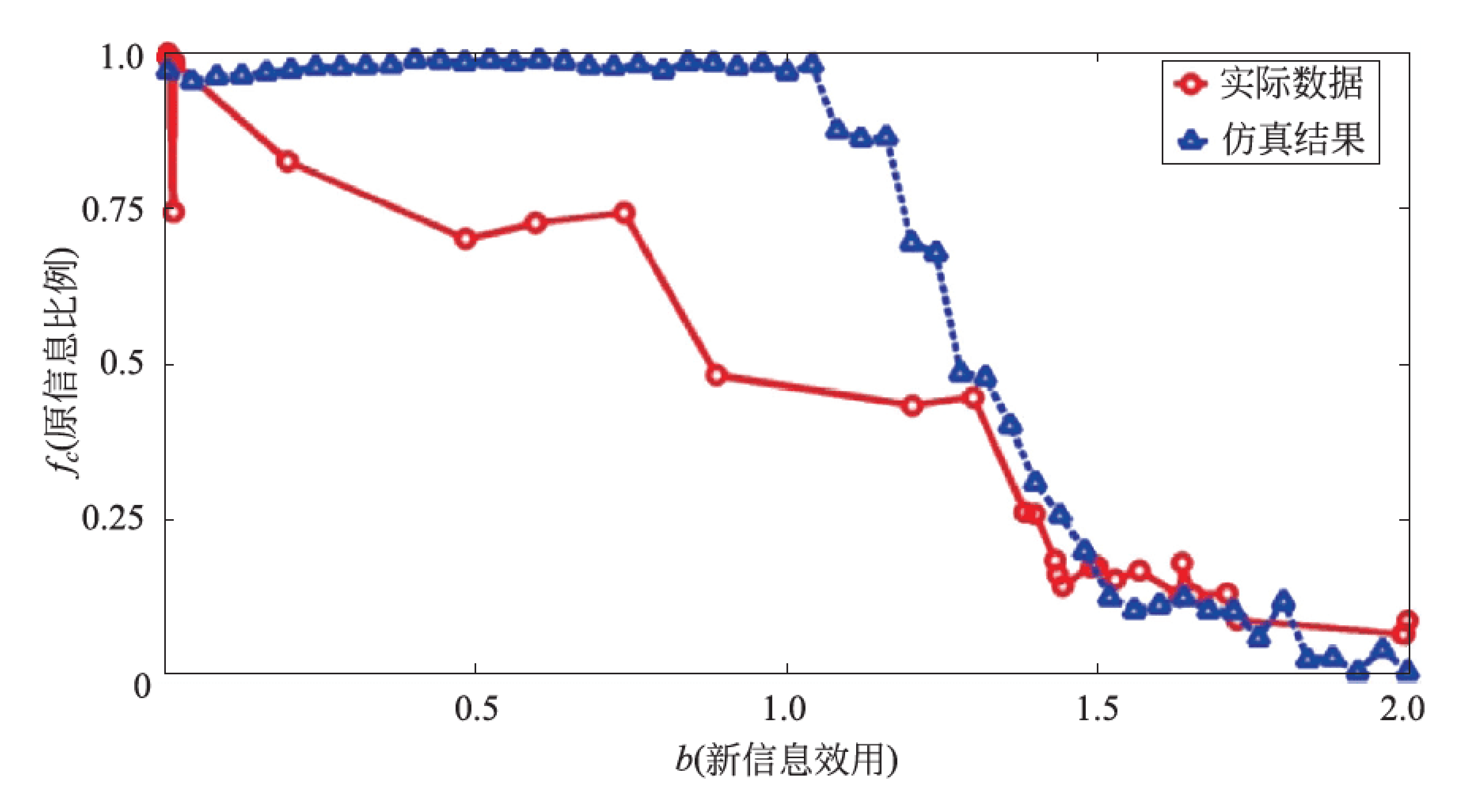


图 12-10 实际数据与仿真结果对比分析($c=0.1$, $W=0.2$, $p_r=0.8$, $ratio=0.5$)

从图 12-10 中可以看出,红色曲线表示实际数据f_c随b的波动情况,蓝色曲线为实验仿真得到的变化趋势,两条曲线有着相似的吻合趋势,由于在实际情况中存在其他电影的上映以及其他一系列不确定因素造成的影响,两条曲线存在一定的差异。当b值小于 1 时,f_c值均大于 0.5,说明微博用户在《大圣归来》影片吸引度不够大时并不会主动、积极地为影片宣传,而是坚持讨论先在微博平台上热议影片《小时代 4》的相关话题。随着b值的逐渐增大,f_c值也随之不断下降。用户开始放弃坚持原有信息,转而选择新加入的新信息。《大圣归来》在微博上的推广营销越发引人关注,微博用户开始主动搜索"大圣归来"相关话题信息,积极推广宣传《大圣归来》,而放弃对原先感兴趣的《小时代 4》相关话题的关注。说明在这次电影微博平台竞争中,虽然《大圣归来》在《小时代 4》之后上映,但是仍然凭借良好的制作与口碑,成功逆袭,在微博传播中抢得先机。可以看出,实际数据曲线与仿真曲线均在$b=1.2$处,f_c值出现下降的趋势。这也很好地验证了声誉因素在竞争性信息传播中所造成的影响,竞争信息必须具有足够高的效用吸引用户,弥补用户切换信息所造成的声誉损失,

这时才能够在信息竞争中占据优势，赢得用户。

12.2.2　新浪微博中电视剧竞争信息异步传播分析

本节使用的数据依然采集自新浪微博。我们从新浪微博平台上采集约 700 000 个用户在 2016 年 4 月 3 日至 5 月 3 日关于《武神赵子龙》和《最好的我们》两部电视剧相关博文的转发、回复以及搜索。这两部电视剧在 2016 年 4 月初先后上映，微博上网友对两部电视剧的热议程度可以视为社交网络上两类信息异步竞争传播这一动态行为。先上映的电视剧《武神赵子龙》可以作为社交网络竞争性信息中的原信息，而随后于 4 月 8 日上映的《最好的我们》可以视为后加入竞争传播的新信息，在原信息开始传播 5 天之后加入竞争。考虑到社交网络中的网络聚集效应：用户群体更愿意把他们的关注点集中在他们所共同的兴趣点上，用户更愿意讨论转发他们所共同关注的电视剧的相关信息。

我们采集两部电视剧相应关键词的原创博文及其转发数，以此作为两部电视剧的热议度来衡量原信息和新信息在社交网络中的影响力和传播范围大小，可以用来描述模型仿真中合作者比例 f_c 这一参数；用户关于两部电视剧关键词的搜索数来衡量原信息和新信息对用户的吸引力，表示模型仿真中收益矩阵的参数 b；两部电视剧官方微博的共同好友数的变化量，用来描述收益矩阵中的参数 c，刻画用户对不同竞争信息的接纳程度。由新浪微博所采集的相关数据已经对数据进行过滤处理，删除了垃圾网站所造成的错误信息，保证了实验数据的真实性与可靠性。

1. 热议度分析

《武神赵子龙》和《最后的我们》分别对应原信息和新信息，其热议度，即相关博文的转发数的变化情况在图 12-11 中显示。红色折线表示《武神赵子龙》原信息的热议度变化趋势，而蓝色折线则代表《最后的我们》的新信息。原信息《武神赵子龙》的热议度占两个信息总热议度的百分比在图 12-12 中表示出来，而仿真模型中 f_c 则可以很好地描述用户坚持选择原信息所占的比率。

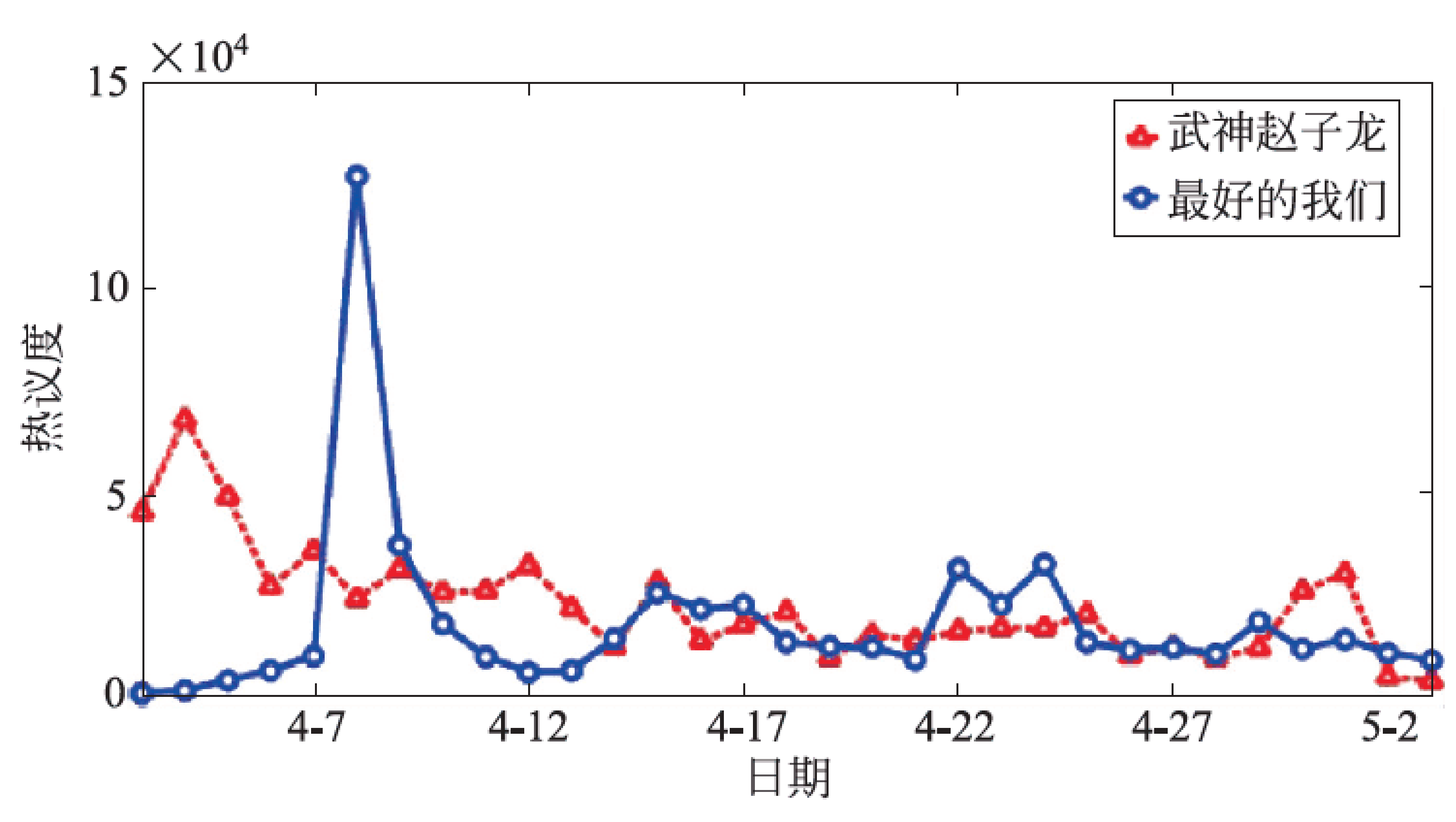


图 12-11　《武神赵子龙》和《最后的我们》热议度（相关博文转发数）统计分析

从上述两张图可以看出，在 2016 年 4 月初，新信息还未开始传播，新浪微博上关于原信息相关博文的转发居多，在两竞争性信息传播中占得先机。而在 4 月 7 日，原信息已传播 5

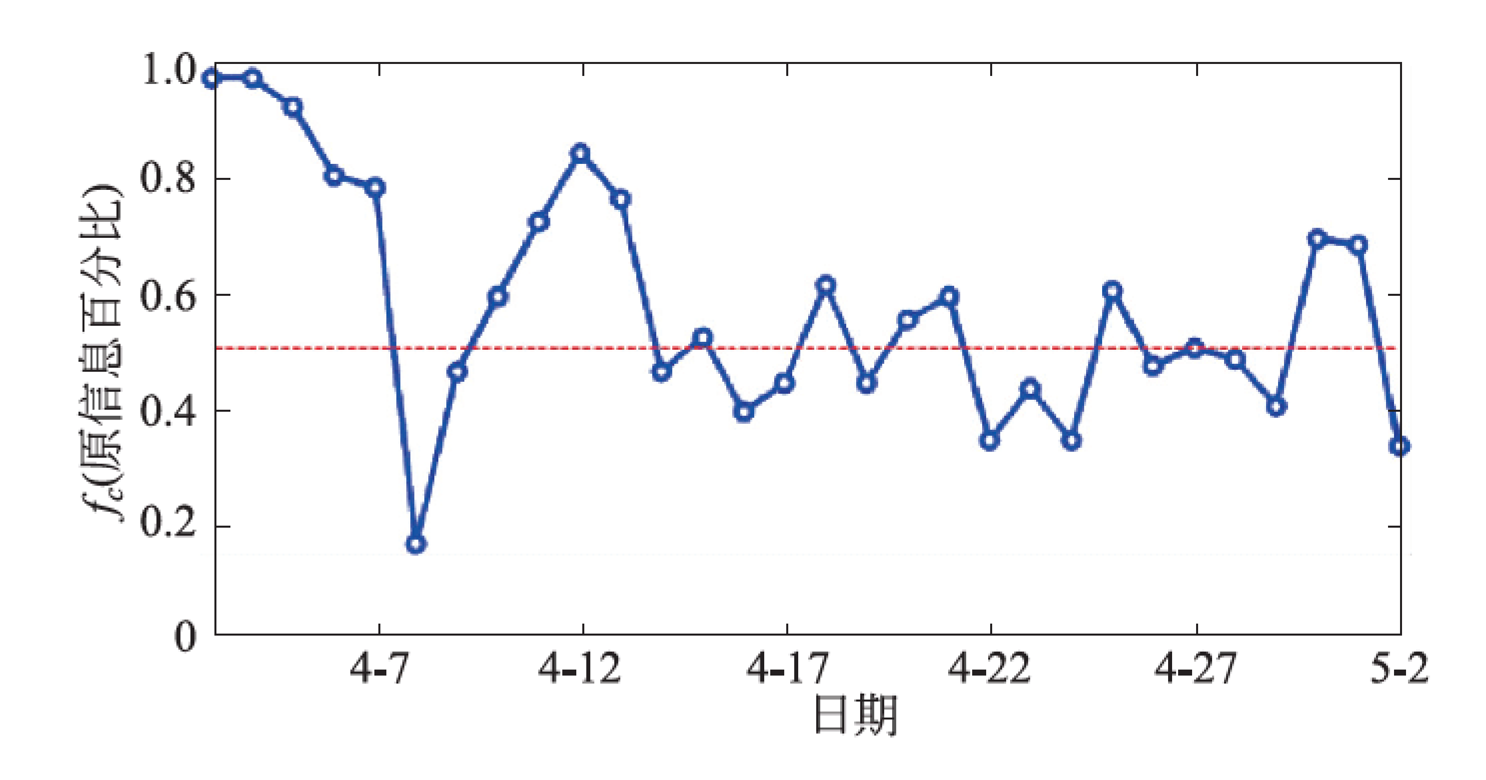


图 12-12　原信息《武神赵子龙》传播百分比

天之后，新信息开始加入竞争，微博上对新信息的转发讨论急剧增大，达到一个峰值，用户的关注度迅速集中在新信息上，新信息的热议程度反超原信息。随着两竞争性信息在微博上的持续传播，4 月中旬～5 月初两竞争性信息的热议程度交替占据优势。原信息与新信息在竞争传播过程中并没有一方完全压制一方，两信息共存演化竞争。

2. 吸引力分析

两部电视剧相关关键词的搜索评论数的多少可以反映两竞争性信息对网络用户吸引力的大小。搜索评论量越大，电视剧越能吸引微博用户的关注，该信息在竞争传播中给用户带来的效用越大。两竞争性信息效用变化情况如图 12-13 所示。红色折线表示《武神赵子龙》原信息效用波动，蓝色折线代表新信息《最好的我们》。

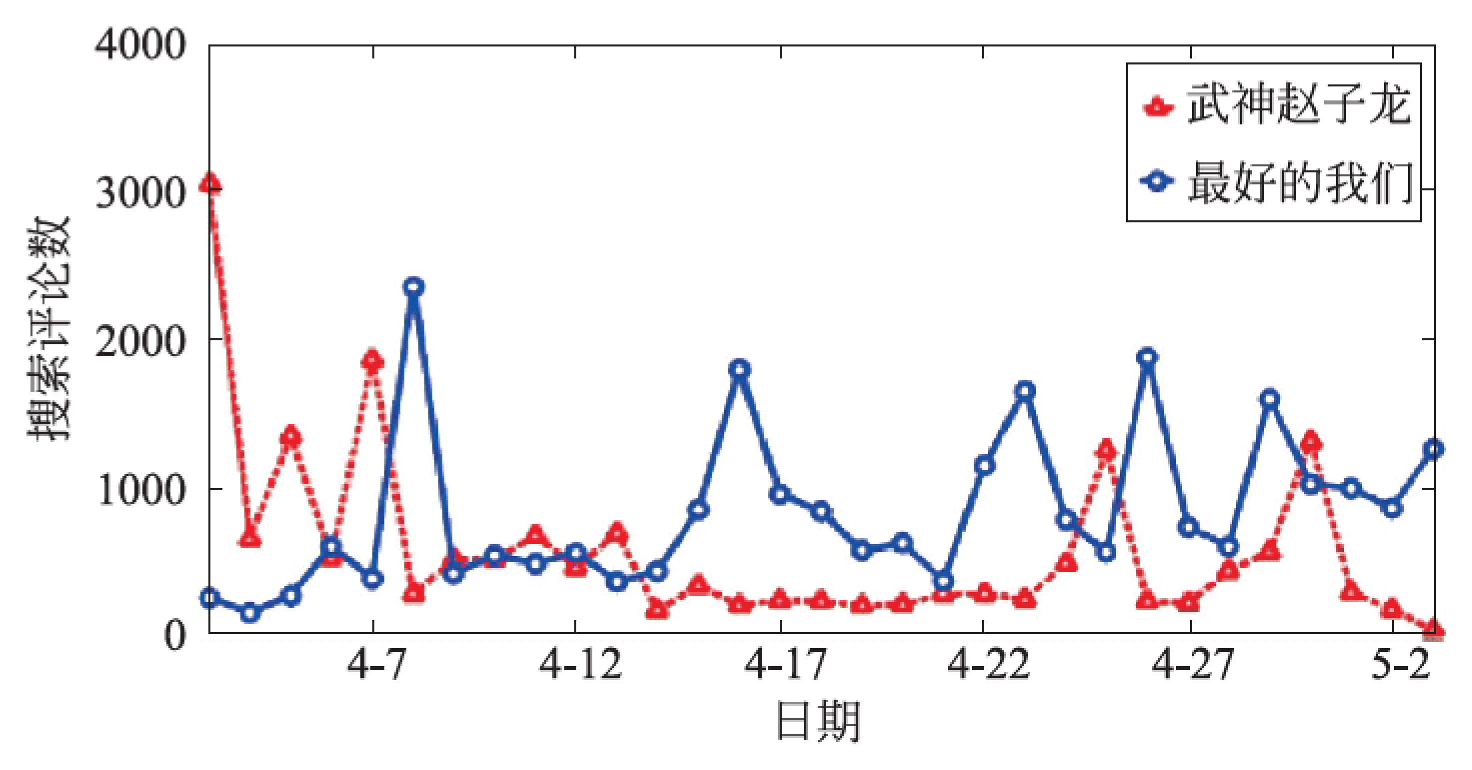


图 12-13　《武神赵子龙》和《最后的我们》效用
（相关关键词搜索评论数）统计分析

图 12-13 显示了 4 月初先传播的原信息相关词条的搜索评论量高于未开始传播的新信息，用户倾向于关注原信息，未加入竞争的新信息对他们的吸引力不够大。当 4 月 7 日新信息开始在微博上传播，微博用户对新加入的信息自然更感兴趣，相关搜索评论数也急剧上升，相应的原信息搜索评论数开始下降，用户的关注点转移到之后加入竞争传播的新信息上。新信息给用户带来的效用超过了原信息的效用。随着两信息竞争性之传播，用户新鲜感消退，两类信息的搜索量不相上下。新信息在传播过程中凭借一系列出众的剧情和优良制作在一定程度上吸引了更多用户的关注，但原信息也依靠一定的营销传播方法赢得了微博用户。新信息带来的效用略高于原信息，但原信息也在某些时间段内占据一定优势。

3. 对比分析

为了能够将实际数据分析结果与模型仿真结果进行对比分析，我们将 2016 年 4 月 3 日至 5 月 3 日《武神赵子龙》和《最后的我们》两部电视剧，即原信息和新信息的热议度（相关博文转发数）与效用（相关关键词搜索评论数）用折线图描绘出来，与模型仿真中 $c=0.1, W=0.2, p_r=0.8, ratio=0.5$ 的仿真结果的折线图进行比较分析。两信息在一个月的时间内先后在新浪微博上的传播过程来验证分析社交网络中异步竞争性信息传播这一现象。相关参数的设定依据文献。原信息效用我们假定为单位 1，后加入竞争的新信息效用 b 设定为以原信息效用单位 1 为标准在 0～2 范围内波动。模型仿真中 f_c 可以用原信息热议度所占总热议度的百分比来描述。f_c 越大，表示原信息在微博上的消息传播范围越广；反之，则表示新信息在信息竞争传播中更具优势。

图 12-14 对比了模型仿真与实际数据波动情况。红色折线反映了真实数据信息竞争传播演化情况，蓝色折线是仿真结果。当 $b<1, f_c>0.5$ 时，说明新信息相关博文对微博用户吸引力不够大，用户不愿意转发其相关博文，更愿意相互之间讨论原信息的相关消息。随着新信息相关博文吸引力的增大，用户开始关注后加入竞争的新信息，并且逐渐引发相互之间的热烈讨论，而对于他们原先所关注的原信息的讨论热情逐渐减弱，f_c 也随着 b 的增大逐渐下降。特别注意到，实际数据折线 f_c 的值也在 1.0～1.2 的区域范围内急剧下降，越过了 0.5 这一临界线，这一现象与我们之前仿真得到的结果相一致，说明声誉因素确实对竞争性信息传播产生影响，使得用户个体不会轻易改变已接受的信息。新信息必须有足够的影响力来吸引用户关注以弥补用户放弃原有竞争性信息的声誉损失，以此来赢得信息竞争传播的胜利。

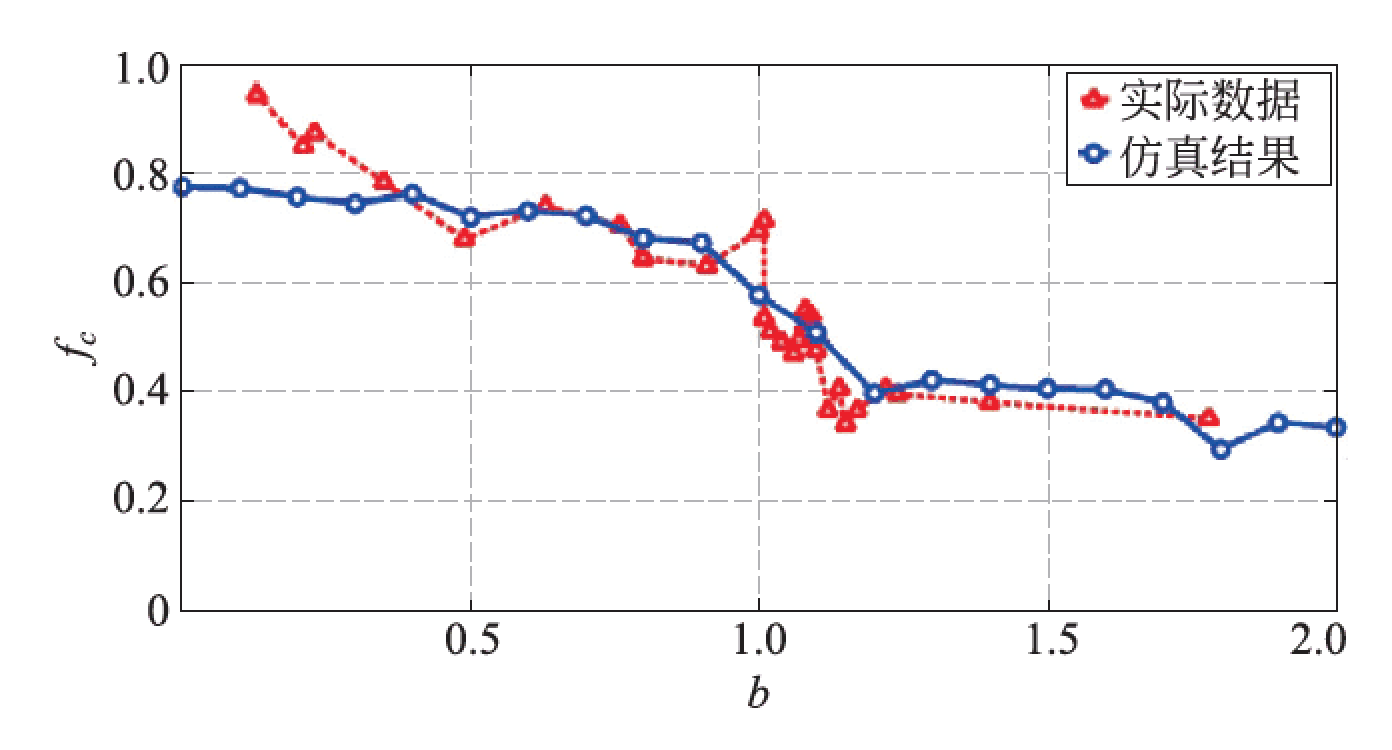


图 12-14 实际数据与仿真结果对比分析
（$c=0.1$，$W=0.2$，$p_r=0.8$，ratio$=0.5$）

两条折线良好的拟合趋势验证了社交演化博弈模型分析异步竞争性信息传播的可行性。由于同期还有其他电视剧的上映，以及其他不确定因素的影响，两条折线仍存在一定的误差，之后的工作将进一步优化模型，调整仿真参数，以期达到更适合描述实际场景的仿真结果。

12.3 新浪微博中电商间信息的竞争性传播分析实例

我们选取 2015 年春节期间新浪微博上广泛流行的“让红包飞”活动作为分析案例，选择了电器行业中具有代表性的两家电商：四季沐歌和美的，将两家电商的微博讨论传播以及

两家电商所派发的红包传播可以看作两个竞争性信息在同一时间上在社交网络上竞争传播的行为活动。该数据集包含了四季沐歌和美的两家公司在 2015 年 2 月 2 日至 2 月 28 日的时间段内相关话题的转发数据。数据集还包含了超过 70 万用户在 2015 年 2 月的状态信息，这些用户构成了一个双向交互的无向网络。四季沐歌的相关话题可以看作竞争传播中的竞争信息 A，相应地，美的的相关话题为竞争信息 B。两家电商所派发的红包数量和相关话题热议程度分别用来刻画两个竞争信息对微博用户的吸引力和信息传播程度。考虑到网络上的焦点效应：网络用户的注意力往往集中在他们所感兴趣的信息资源上，表现在微博用户更喜欢共同分享同一家公司所派发的红包。这一网络特征现象可以很好地由协调博弈策略来表示。

在本节中，我们分析四季沐歌和美的两家电商相关话题的传播过程和特征，并且讨论了竞争信息在微博社交网络平台上的演化情况，并与 5.2.8.2 节中的仿真结果进行了比较。

1. 热议度分析

根据四季沐歌和美的原创微博和转发微博数据统计得出两个电商相关关键词热议度的变化趋势。关键词的热议度反映了相关电商在网络用户中受欢迎的程度和相关信息在微博中的传播范围，对应描述模型仿真中平均合作率 f_c。在实际数据中，平均合作率 f_c 具体反映出四季沐歌相关关键词信息的热议度在两者总电商总热议度中所占的百分比。

从图 12-15 中可以看出，四季沐歌和美的相关关键词热议度曲线均呈现出一定的波动趋势。图中蓝色实线表示四季沐歌热议度变化情况，而红色虚线则代表美的热议度。两家电商的热议度交替占据上风，并在某个特定的时间段内达到一个峰值。两家公司都采取了一系列行之有效的营销方法，并在“让红包飞”活动中大量分发企业红包，吸引了众多微博用户的关注。在一些特殊的日子，比如春节除夕前夕，四季沐歌和美的大幅度提升派发红包的数目和红包金额，短时间内引起众多网友的关注，在微博用户中引发相关主题微博的热烈讨论的海量转发。更值得注意的是，相对于美的热议度在一定范围内相对平稳的波动，四季沐

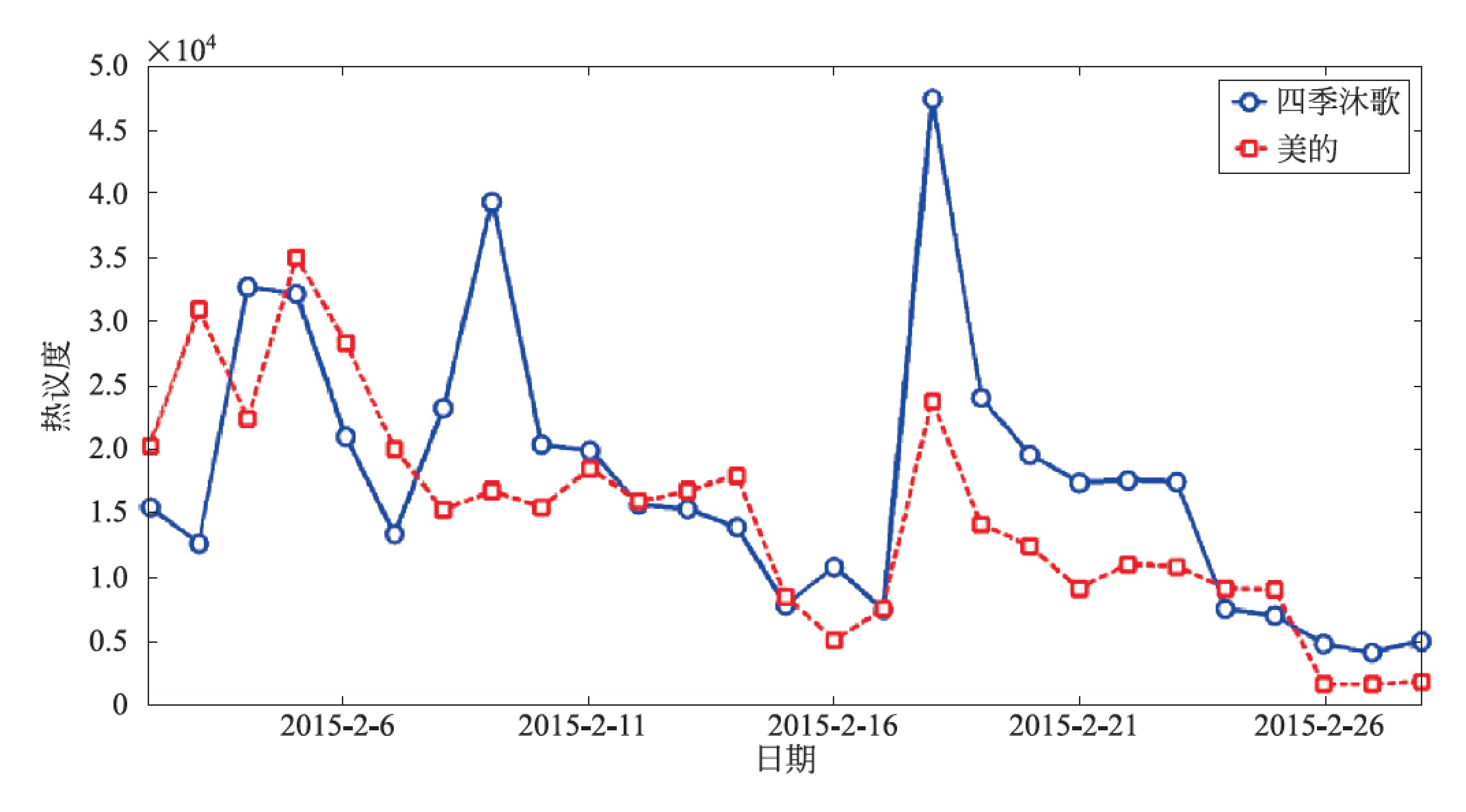


图 12-15 “美的”和“四季沐歌”热议度(相关博文转发数)统计分析

歌热议度折线的变化趋势更为激烈，明显地呈现出 3 个较大的峰值。在这段时间内，四季沐歌通过一定的网络营销推广，使得相关微博话题迅速传播开来，成为微博用户转发讨论的焦点。

在图 12-16 中，f_c 反映出四季沐歌相关关键词微博在与美的相关微博竞争传播中热议度所占百分比。虚线标识出 $f_c=0.5$ 的分界线，表示四季沐歌和美的相关微博热议度各占 50%，两个电商相关信息在竞争传播中平分秋色。而实线代表四季沐歌热议度百分比 f_c 的变化情况。我们将四季沐歌相关关键词微博看作是信息 A，美的相关关键词微博看作是信息 B。当 $f_c>0.5$ 时，呈现出实线在虚线上方波动，用户更偏向转发、分享信息 A，并且更坚持选择他们已经有所偏好的微博进行分享，这个时间段内，四季沐歌在竞争传播中占据优势。相反地，当 $f_c<0.5$ 时，用户开始放弃信息 A，转而选择分享、转发信息 B，美的的相关微博在微博用户中流行得更为广泛。从折线的整体趋势可以看出，四季沐歌相关话题在竞争传播中占据了一定优势，但是美的也不甘示弱，特别是在月初的竞争期间，抢走了相当一部分的市场份额。但是四季沐歌很快通过派发大量红包以及各种移动终端互动推广活动，很快得到用户的强烈关注，在信息竞争中扭转了劣势，占据信息传播的主动权。在月末“让红包飞”活动的尾声，尽管美的增大了营销力度，丰富了促销推广活动形式，并且在一段时间内在与四季沐歌的竞争中重新取得优势，但就整体而言，四季沐歌仍然具有在这次信息竞争中的优势。“让红包飞”活动中四季沐歌和美的相关微博话题的传播可以看作两个竞争信息同时传播争夺用户的行为活动，信息本身的效用变化引起用户对其关注。接下来，我们从电商派发红包数目来分析信息效用的变化情况。

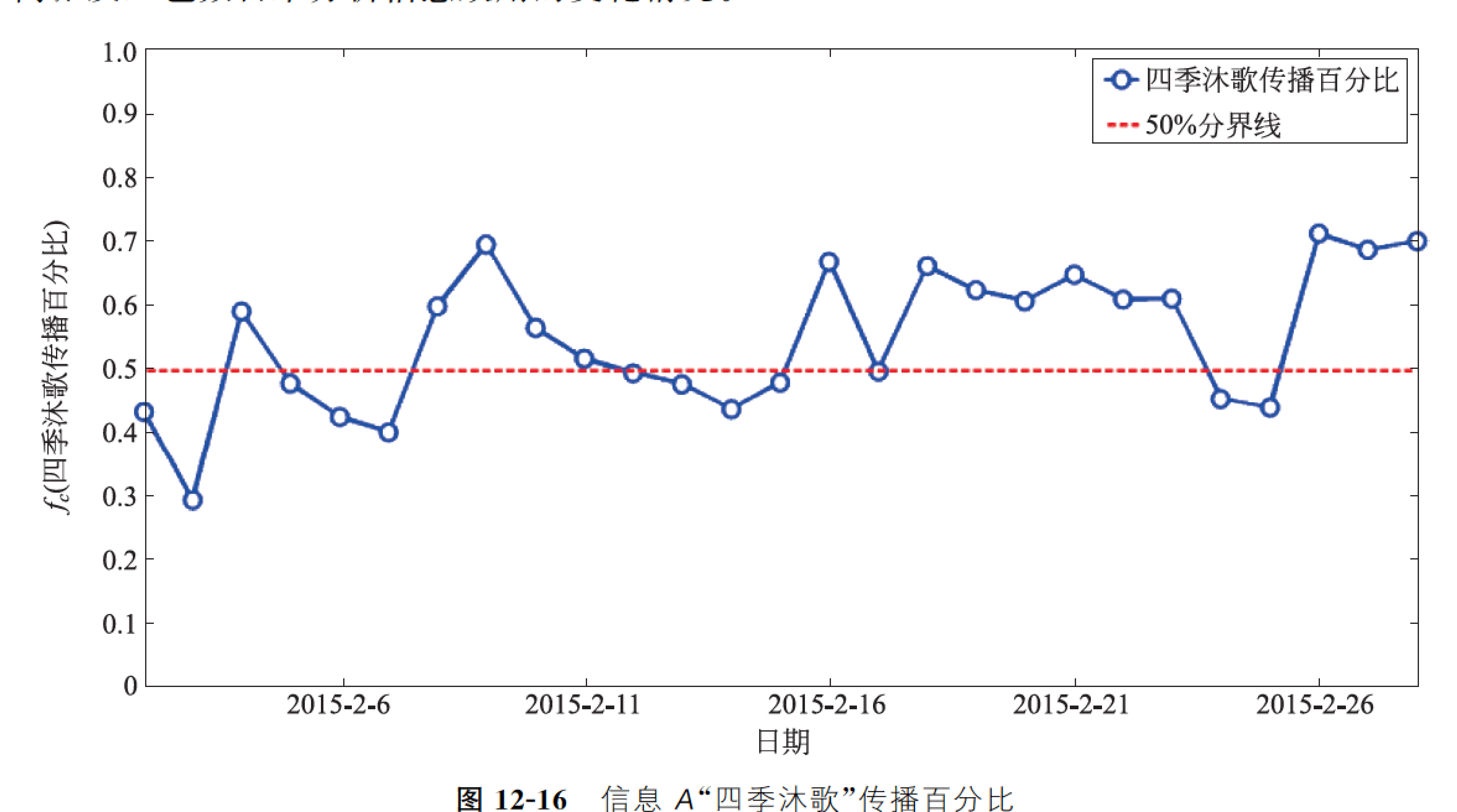


图 12-16 信息 A“四季沐歌”传播百分比

2. 吸引力分析

我们采集用户对于“四季沐歌”和“美的”两个关键词的搜索数和两个电商在 2015 年 2 月“让红包飞”活动中所派发的红包数量，以此来作为效用参数反映两个竞争信息对用户吸引力的大小。四季沐歌的效用我们设置为单位 1，以四季沐歌效用为标准，美的的效用我们

用 b 来表示。通过标准化处理，b 的取值范围为 0～2。

图 12-17 显示了两家电商相关话题信息效用的变化情况。虚线代表作为单位 1 标准的四季沐歌的效用，围绕单位 1 上下波动的实线代表美的相关话题效用。在 2 月初期，美的的效用更多地超过单位 1 四季沐歌效用，用户更多地主动搜索美的的相关话题微博，更多地分析其派发的红包。随着时间的推移，美的相关话题效用逐渐下降，其实线波动到单位 1 虚线的下方。四季沐歌的相关话题推广开始发力，用户开始广泛搜索四季沐歌的相关话题微博，更多地去选择转发分享其分发的红包。2 月中下旬的春节期间，四季沐歌的相关话题搜索数更占据一定优势，但是美的通过派发红包，推广营销也成功增加了用户对其的关注度和搜索量。两家电商在微博上所派发的红包数量很大程度上影响了用户对其相关话题微博转发的主动性。

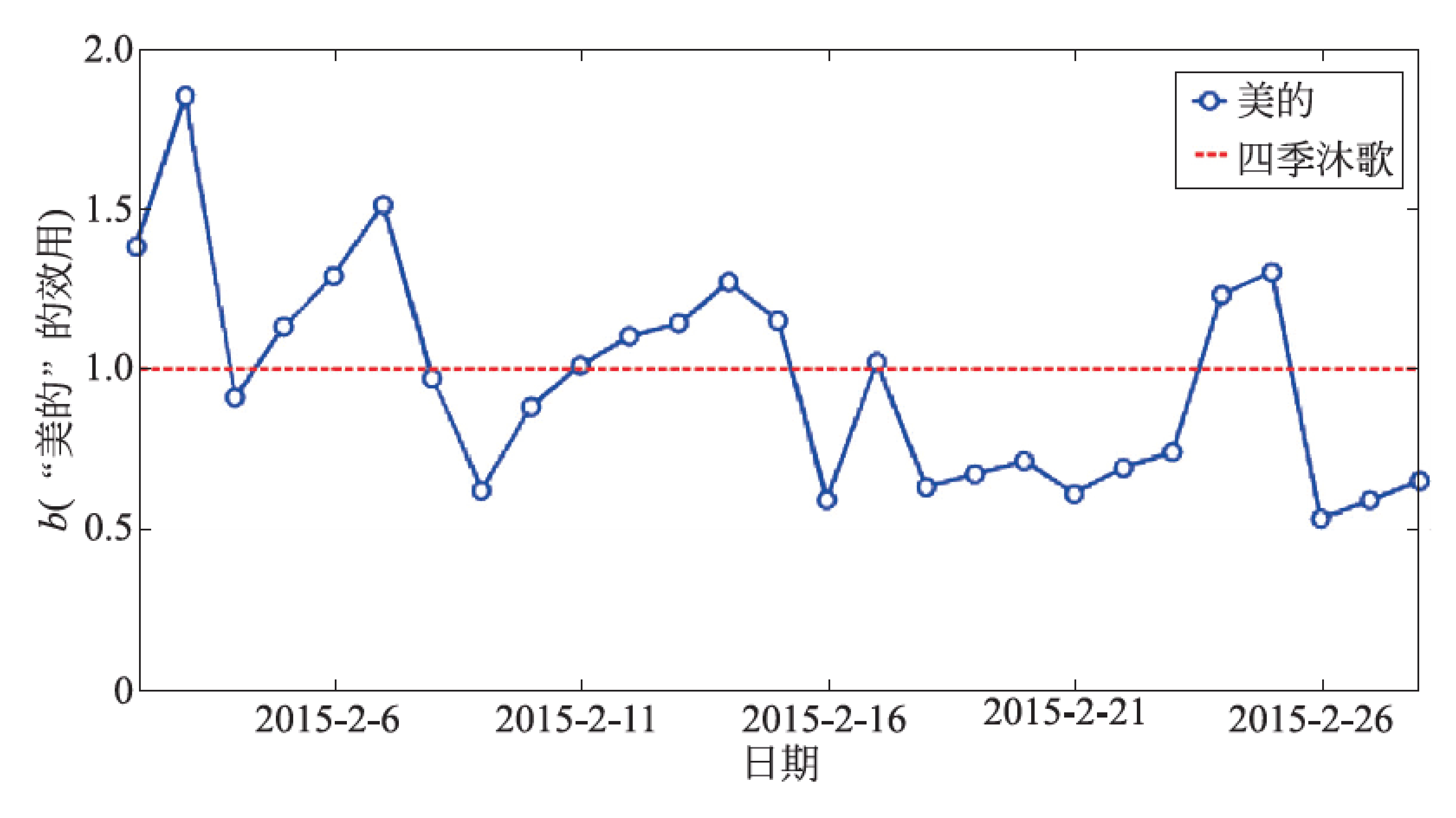


图 12-17 "四季沐歌"和"美的"效用(相关关键词搜索评论数)统计分析

3. 抢红包对比分析

为了将竞争传播真实数据分析结果与模型仿真结果进行对比验证，我们将平均合作率 f_c 随效用 b 的变化情况在图 12-18 中呈现出来，并且与前面章节中图 5-23 中较低水平 p_r 环境下的仿真演化结果进行比较分析。模型仿真验证数据我们选择 $p_r=0.3$ 和 $W=0.8$ 参数下效用 b 和平均合作率 f_c 的仿真结果，并与"让红包飞"活动中两个竞争信息传播演化数据进行验证对比。

虚线表示真实场景中实际数据 f_c 随 b 的变化情况，实线表示对应仿真数据。由于在"让红包飞"活动中还有其他许多企业也参加了推广活动，同样派发了大量的红包，吸引了用户的关注并且在实际环境中还存在其他一系列的不确定因素，导致两条折线仍有一定的误差。当 $b<1$ 时，$f_c>0.5$，倘若美的所派发的红包对用户的诱惑力并不足够大，用户并不会主动选择分享其相关话题信息。用户这时会选择坚持转发四季沐歌的相关话题，并分享其企业红包。f_c 随着 b 的增大而减小，意味着用户开始放弃四季沐歌的相关竞争信息 A，转而选择转发美的竞争信息 B。良好的市场营销给美的带来了微博用户关注，吸引用户主动搜索其相关话题微博。用户对企业电商是否关注，对其相关话题微博转发是否具有主动性，很大程度上与企业所派发的红包量相关。

实际数据折线和模型仿真折线的 f_c 均在 $b=1.2$ 左右呈现出明显的下降趋势。这一现

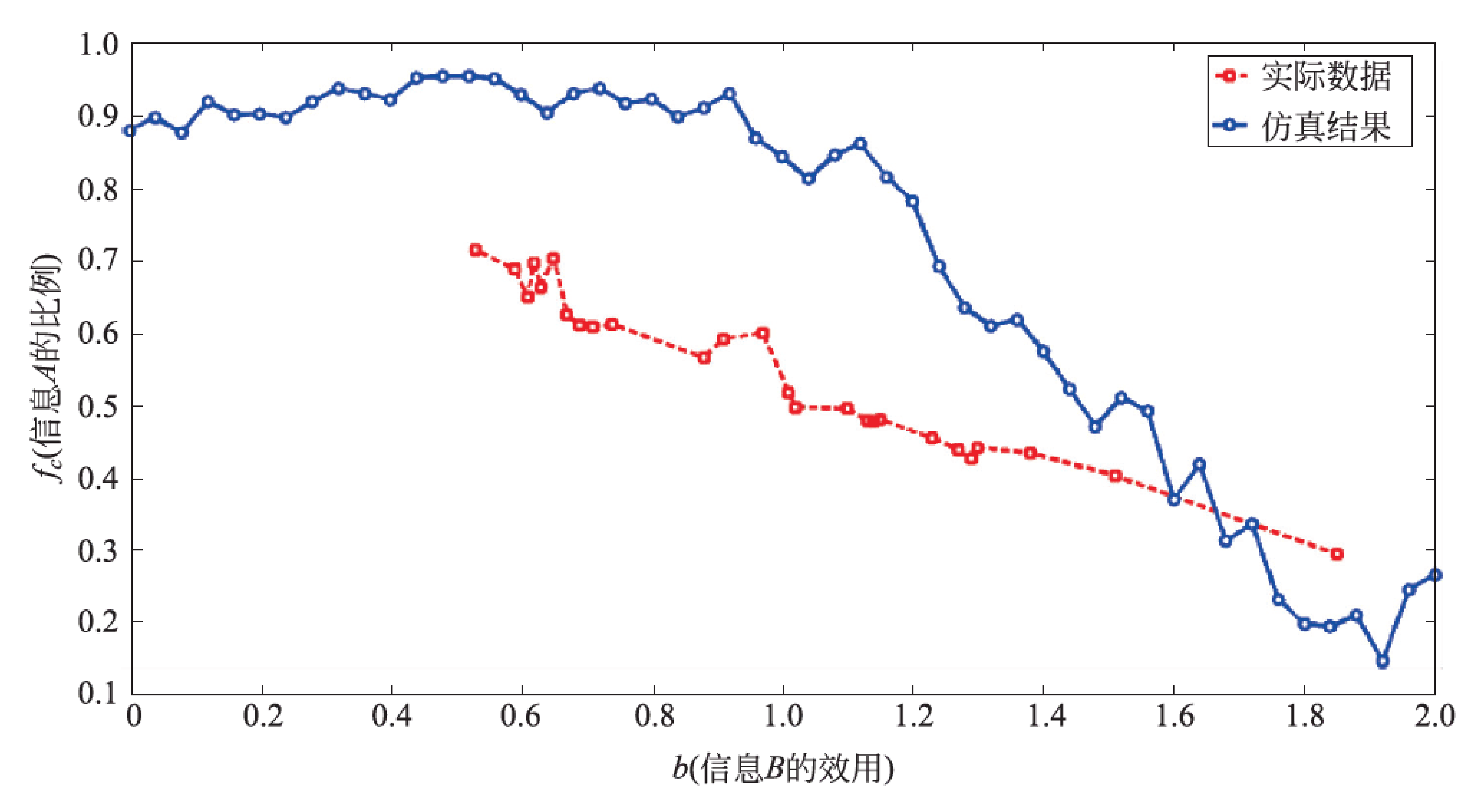


图 12-18　实际数据与仿真结果对比分析（$W=0.8$，$p_r=0.3$）

象很好地验证了竞争性信息传播中声誉因素的影响。总之，竞争信息对用户的吸引力必须足够大，足以弥补用户改变信息选择时造成的声誉损失，只有这样，用户才会选择改变信息选择，相关竞争信息才会在传播中占据优势。

12.4　足球圈在新浪微博和腾讯微博的使用热度趋势分析实例

当社交网络中的用户数量迫近世界网民总量的时候，获取新用户或者榨取老用户的更多使用时间变得越来越困难。我们可以形象地将该发展阶段命名为社交网络服务的发达阶段。当前，在线社交网络已经进入了发达阶段，新的用户资源变得越来越稀缺，社交网络运营商被迫要与其他社交网络运营商竞争，从其他运营商处获取更多的用户关注度，从而保证自己在行业中的地位。

本节对国内用户量和用户活跃度最高的新浪微博与腾讯微博的真实数据进行对比，进一步直观地给出两个社交网络之间的用户关注度博弈过程。我们选取的是在中超联赛中同时拥有新浪微博与腾讯微博账户的足球运动员作为实验对象，共计 107 人。其中，“辽宁宏运”和“长春亚泰”在 2011 年与腾讯微博签署了推广协议，这两队共 24 人在腾讯微博的活跃度会受此影响，我们将其定义为推广账户，并与普通账户进行了对比。

选用中超联赛运动员作为实验对象的理由在于：①足球运动员之间联络比较紧密，互动较多；②不同于演艺圈等名人账号存在团队经营的情况，足球运动员的账号以个人直接使用为主，这更能体现出个体的演化博弈过程。

我们以一个用户在新浪微博和腾讯微博发布的消息之和作为该用户的全部关注度，将过去 5 年间的消息分别进行关注度统计。图 12-19 为足球圈平均的关注度，图 12-20 为足球圈推广账户的关注度，可以看出，自 2011 年起至 2015 年止的这 5 年中，新浪微博的用户关注度始终领先于腾讯微博。进一步比较图 12-19 和图 12-20 可以发现，被腾讯微博推广策略影响的账户，其最终在腾讯微博的活跃度却在 2015 年彻底消亡了。这一方面说明，对单独的小群体的推广很难扭转新浪微博不断扩大领先优势的大趋势，另一方面，也证明了邻居

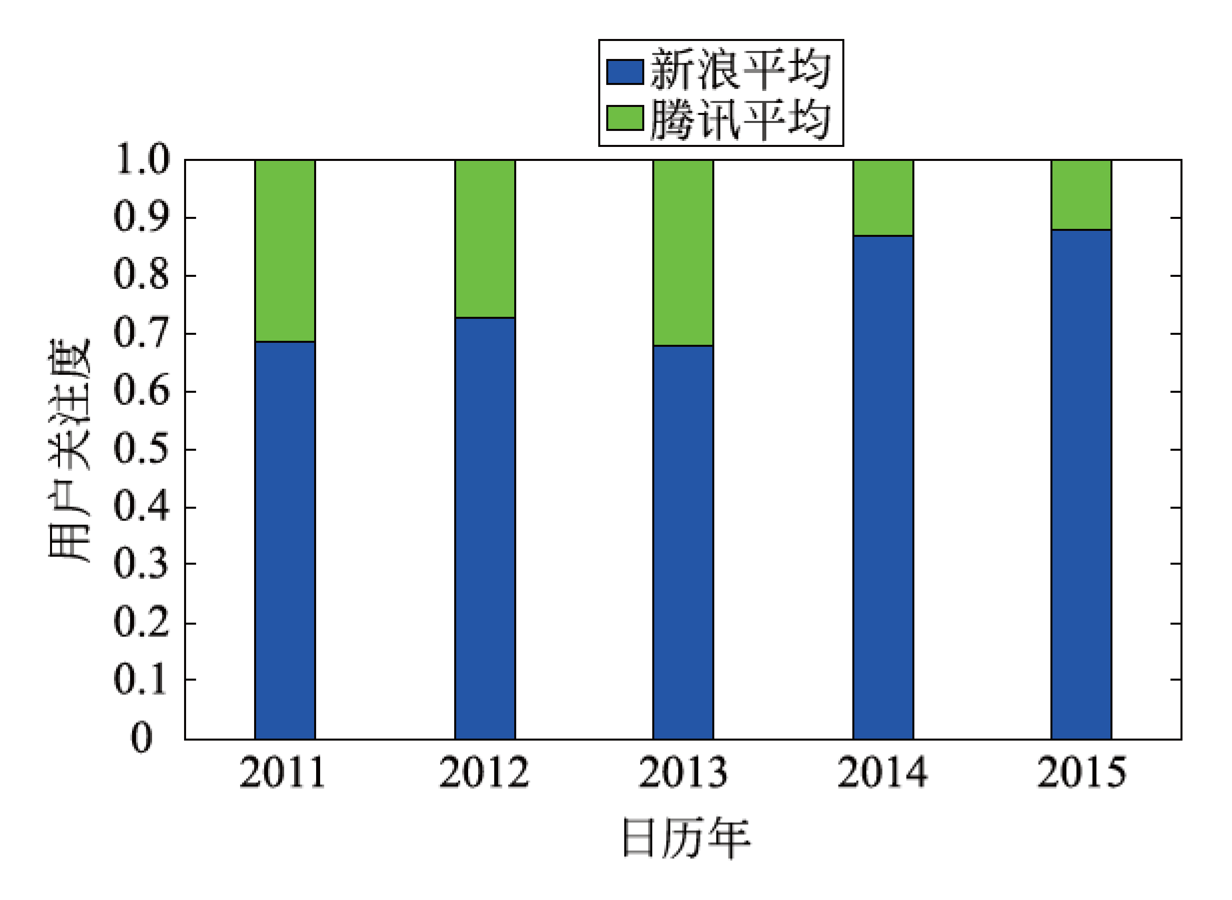


图 12-19 足球圈平均关注度对比

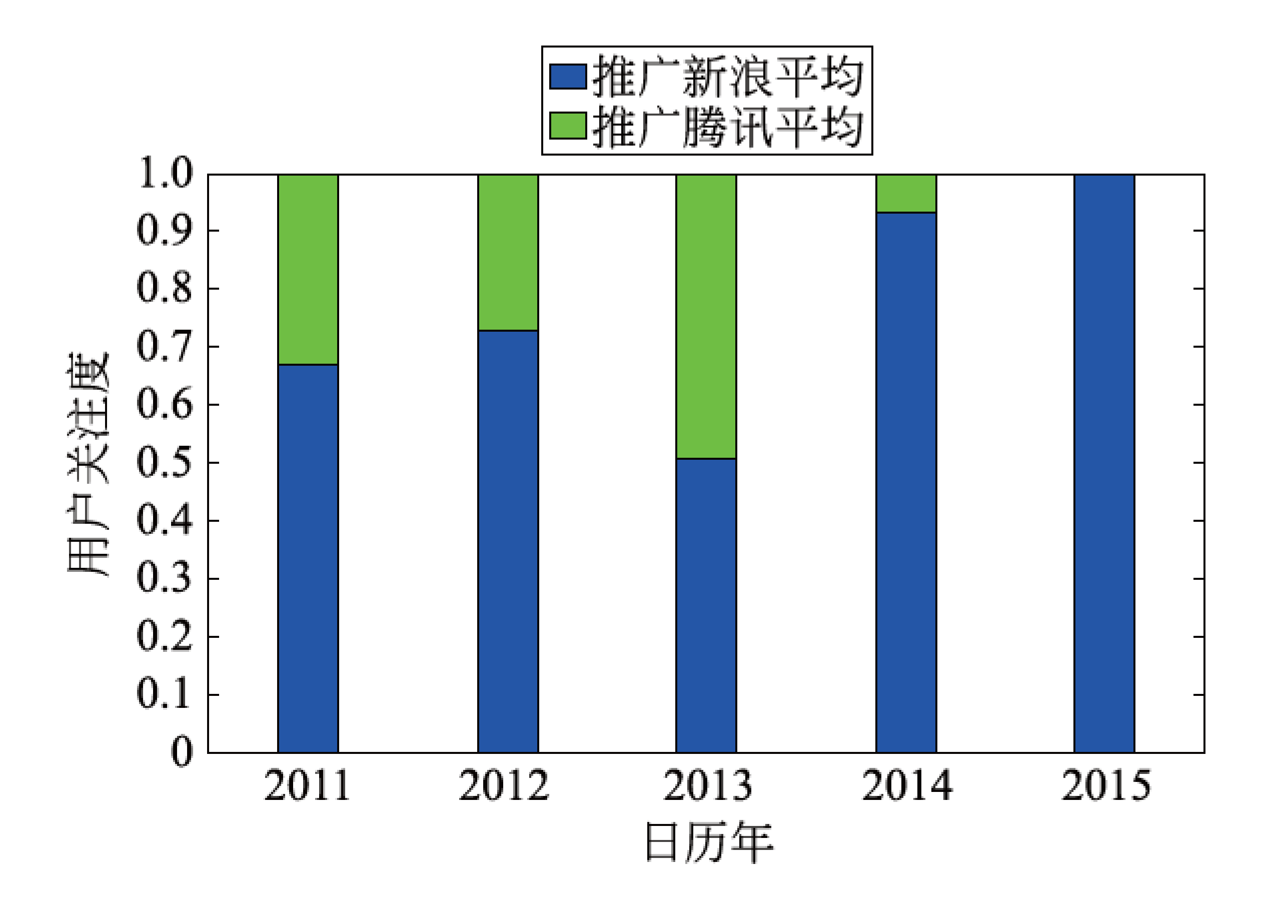


图 12-20 足球圈推广账户平均关注度对比

间在不同社交网络上关注度的变化确实可以影响并且加速整个群体关注度由一个社交网络向另一个社交网络的转移。

在分析一个群体的用户关注度分布,特别是如微博这类社交网络的用户关注度分布时,由于用户的活跃度服从长尾分布,其关注度的中值往往更能反映出群体的真实关注倾向。图 11-21 为整个足球圈的用户关注度中值与推广账户的用户关注度中值的对比图,从图中我们可以看出:①足球圈用户中近 5 年间新浪微博与腾讯微博的关注度从约 50%~50%分

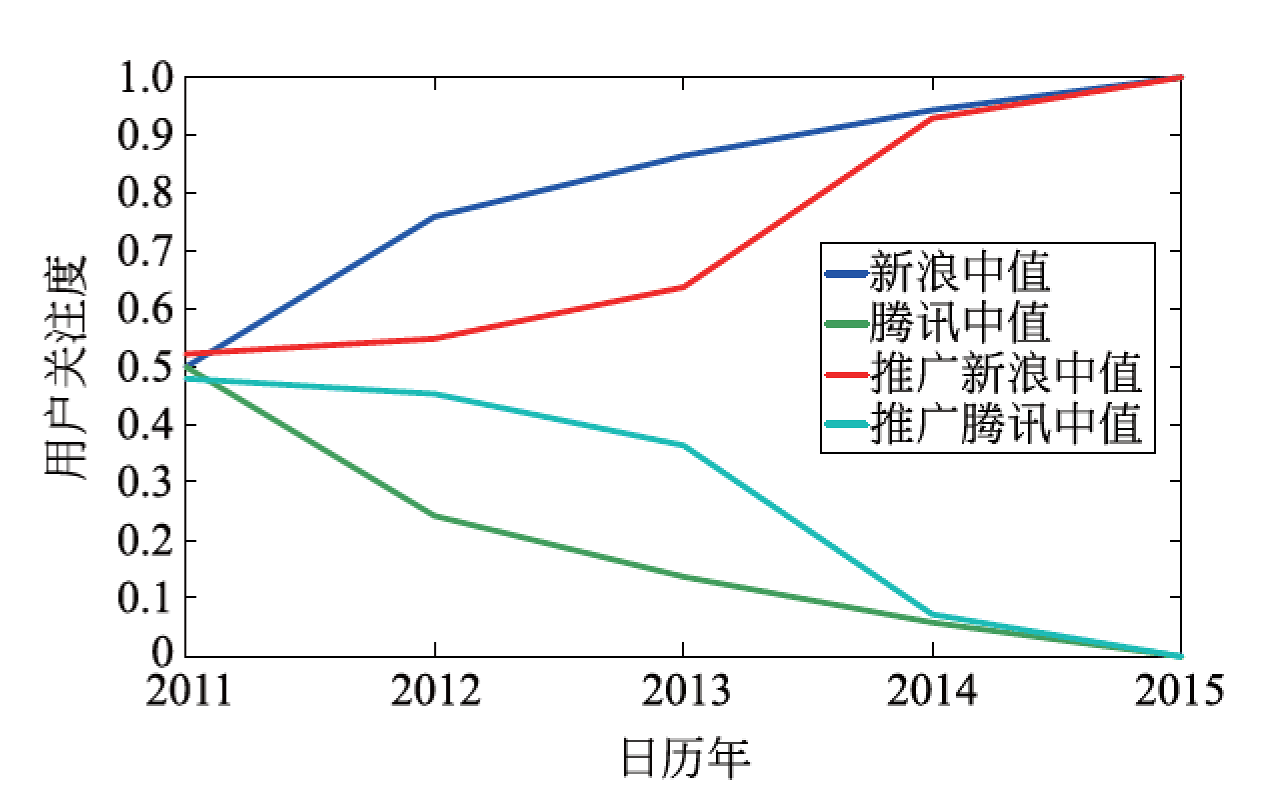


图 12-21 足球圈中值与推广账户中值对比

布，逐渐形成了新浪微博彻底压倒腾讯微博的局面；②腾讯的推广确实对腾讯微博的关注度下降起到了延缓的作用，在 2013 年，中值与推广中值之间的关注度差异甚至超过了 0.2，但最终仍无法阻止腾讯微博整体式微。

小结

本章通过对社交演化博弈应用的实际案例分析来说明社交演化博弈模型的应用效果。

通过对新浪微博和 Twitter 的实际数据的采集和分析，分别估算了两个社交网络的正向行为比例、网络聚集系数和幂律指数，并与社交网络演化博弈模型仿真结果进行对比分析，发现社交演化博弈模型对于社交网络演化的仿真与真实世界中的社交网络的演化发展在一定程度上是吻合的，验证了模型的有效性。

对于竞争性信息异步的传播，根据竞争信息影响持久度，选取了两类异步信息竞争案例：电影类和电视剧类。通过对新浪微博上热映电影和电视剧的相关微博搜索转发量的实际数据的统计分析，并结合社交演化博弈模型仿真结果进行对比分析，说明了声誉在竞争性信息异步传播中的影响力。

对于竞争性信息同步传播，选取了新浪微博中“让红包飞”活动中相关微博搜索转发的真实数据，分析了相关因素对竞争传播的影响，竞争信息必须额外提供收益以弥补用户声誉损失，从而吸引用户改变信息选择。

对于微博间用户使用热度的竞争性分析，选取新浪微博和腾讯微博上足球运动员账户群体，给出了其使用变化情况，并给出了分析结论。

参考文献

[1] Yu J, Wang Y, Jin X, *et al*. Social Evolutionary Games. GameNets 2014.

[2] Yu J, Wang Y, Jin X, *et al*. Evolutionary analysis on online social networks using a social evolutionary game. WWW 2014.

[3] Lu Y, Wang Y, Yu J, *et al*. Mechanism analysis of competitive information synchronous dissemination in social networks. In: Proc. of the APWEB. 2016.

[4] 陆源，王元卓，李静远，靳小龙. 社交网络中竞争性信息异步传播模型. 第 4 届 CCF 大数据学术会议. 2015.

[5] 李静远，王元卓，于建业，邢国亮，靳小龙. 基于社交演化博弈的社交网络用户关注度竞争分析. 第 4 届 CCF 大数据学术会议. 2015.